An Information Geometric Approach to Polynomial-time Interior-point Algorithms
— Complexity Bound via Curvature Integral —

Atsumi Ohara [*]   Takashi Tsuchiya[†]

December 2007 (Revised: March 2009)

Abstract

In this paper, we study polynomial-time interior-point algorithms in view of information geometry. Information geometry is a differential geometric framework which has been successfully applied to statistics, learning theory, signal processing etc. We consider information geometric structure for conic linear programs introduced by self-concordant barrier functions, and develop a precise iteration-complexity estimate of the polynomial-time interior-point algorithm based on an integral of (embedding) curvature of the central trajectory in a rigorous differential geometrical sense. We further study implication of the theory applied to classical linear programming, and establish a surprising link to the strong "primal-dual curvature" integral bound established by Monteiro and Tsuchiya, which is based on the work of Vavasis and Ye of the layered-step interior-point algorithm. By using these results, we can show that the total embedding curvature of the central trajectory, i.e., the aforementioned integral over the whole central trajectory, is bounded by $O(n^{3.5}\log(\bar{\chi}_A^* + n))$ where $\bar{\chi}_A^*$ is a condition number of the coefficient matrix A and n is the number of nonnegative variables. In particular, the integral is bounded by $O(n^{4.5}m)$ for combinatorial linear programs including network flow problems where m is the number of constraints. We also provide a complete differential-geometric characterization of the primal-dual curvature in the primal-dual algorithm. Finally, in view of this integral bound, we observe that the primal (or dual) interior-point algorithm requires fewer number of iterations than the primal-dual interior-point algorithm at least in the case of linear programming.

Key words: interior-point methods, information geometry, polynomial-time algorithm, linear programming, semidefinite programming, embedding curvature, computational complexity, differential geometry, convex programming

1 Introduction

1.1 Setting and background

Let $\mathbf{E}$ be an n-dimensional vector space, let $\Omega \subset \mathbf{E}$ be a proper open convex cone, and let $\mathbf{E}^*$ be the space of linear functional on $\mathbf{E}$. We denote by $\langle \cdot\, , \ \cdot \rangle$ the duality product between elements in $\mathbf{E}$

[*]Graduate School of Engineering Science, Osaka University, 1-3 Machikane-yama, Toyonaka, Osaka 560-8531, Japan. (Email: ohara@sys.es.osaka-u.ac.jp).

[†]The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku, Tokyo, 106-8569, Japan. (Email: tsuchiya@sun312.ism.ac.jp). This author is supported in part with Grant-in-Aid for Scientific Research (B), 2007, 19360192 from the Japan Society for the Promotion of Sciences.

and $\mathbf{E}^*$. Let $\Omega^* \subset \mathbf{E}^*$ be the open dual cone $\{s \in \mathbf{E}^* \mid \langle s, x\rangle > 0, \ \forall x \in \mathrm{cl}(\Omega)\backslash\{0\}\}$ of Ω. Associated with Ω and Ω^*, we consider the following dual pair of convex linear programs:

$$\min \ \langle c, x\rangle, \quad \text{s.t. } x \subset (d + \mathbf{T}) \cap \mathrm{cl}(\Omega) \tag{1}$$

and

$$\min \ \langle d, s\rangle, \quad \text{s.t. } s \in (c + \mathbf{T}^*) \cap \mathrm{cl}(\Omega^*), \tag{2}$$

where $c \in \mathbf{E}^*$ and $d \in \mathbf{E}$, $\mathbf{T} \subset \mathbf{E}$ is $(n-m)$-dimensional linear subspace, and $\mathbf{T}^* \subset \mathbf{E}^*$ is m-dimensional linear subspace which are mutually "orthogonal", i.e., for any $s \in \mathbf{T}$ and $x \in \mathbf{T}^*$, we have $\langle x, s\rangle = 0$.

This is a generic framework of conic linear programming which includes classical linear programming established by Dantzig [12], and semidefinite programming by Nesterov, Nemirovski [42] and Alizadeh [1]. Linear programming is widely used as one of the major models in various areas of engineering and information processing, surviving over 60 years. Now we can even solve a linear program with one billion variables and three hundred million constraints [22]. Semidefinite programming has been extensively studied in the last decade with the development of efficient polynomial-time interior-point algorithms and due to its wide applicability in computer science, control theory, statistics, learning theory and pattern recognition etc. [10, 72]. The diverse development of interior-point algorithms was initiated by the seminal paper by Karmarkar who proposed the projective-scaling method for classical linear programming [27].

The purpose of this paper is to establish a direct connection between the complexity theory of polynomial-time interior-point algorithms for conic convex programs and information geometry. These two theories, the former established by Nesterov and Nemirovski [42] the latter by Amari and Nagaoka [3, 4, 6, 41, 51], are among several major innovative developments in computational mathematics and mathematical informatics in the last two decades.

We will demonstrate that information geometry provides a quite suitable differential geometrical framework for analyzing polynomial-time interior-point algorithms for convex programming. Information geometry is a differential geometric framework for informatics and has been successfully applied to several areas such as statistics, signal processing and machine learning. From mathematical point of view, information geometry is nothing but *differential geometry of convexity*. Recently, importance of convexity in the context of modeling is recognized more and more due to the success of convex modeling in various areas of engineering and mathematical sciences based on convex optimization [10, 11, 40]. Therefore, it would be nice if we could develop a theory which tightly connect information geometry and convex optimization, the two major areas where convexity plays fundamental roles, in particular from the viewpoint of complexity.

Based on this perspective, we will develop a polynomial-time path-following interior-point algorithm for conic convex programming based on information geometry. The iteration-complexity of the path-following algorithm is estimated precisely with an integral of (directional) embedding curvature of the central trajectory in a rigorous differential geometrical sense.

We further study implication of the theory when applied to classical linear programming, and establish a surprising link to the strong curvature integral bound for primal-dual central trajectory established by Monteiro and Tsuchiya [39], which is based on the work of Vavasis and Ye of the layered-step interior-point algorithm [71]. The Vavasis-Ye algorithm is a powerful algorithm which becomes a strongly polynomial-time interior-point algorithms for combinatorial linear programs. This direction of research is somehow connected to the problem of existence of a strongly polynomial-time algorithm for linear programming, one of the fundamental open problems in computational complexity theory and computational mathematics [61].

In the following, we will provide brief introductory descriptions on the topics we will deal with.

[Differential geometric approach to interior-point algorithms, mainly based on Riemannian geometry]

So far, there has been several papers which attempted to establish connections between differential geometry and interior-point algorithms. The approach attracts us from the early days of interior-point algorithms when people find that a path called central trajectory plays the fundamental role [23, 32, 49, 52]. One of the first literatures which touches on geometrical structure of the interior-point algorithms is Bayer and Lagarias [8, 9]. Karmarkar [28] suggested an interesting idea of introducing analyzing the interior-point algorithm based on Riemannian geometry and the central trajectory curvature integral. Nesterov and Nemirovski [42] introduced the general theory of polynomial-time interior-point algorithms based on the self-concordant barrier function. After this seminal work, geometrical study of interior-point method is conducted by several authors including [17, 44, 43] in the context of general convex programs, where structure of geodesic and Riemannian curvature etc. are studied based on the Riemannian geometrical structure defined by taking the Hessian of the self-concordant barrier function as the Riemannian metric. While the aforementioned Hessian structure sounds natural in geometrical approach to interior-point algorithm, a difficulty arises when we try to find a direct link between the Riemannian curvature of the central trajectories and the iteration-complexity of the interior-point algorithms.

Another meaningful geometrical machinery for development of polynomial-time interior-point algorithms is symmetric cones and Euclidean Jordan algebra [18]. They provide a transparent unified framework of treating polynomial-time primal and primal-dual interior-point algorithms for symmetric cone programming, the case when Ω in (1) is a symmetric cone [19, 20, 21]. Symmetric cone programming [10] is an important subclass of conic linear programming which includes linear programming and semidefinite programming as special case. It also contains second-order cone programming [2, 66, 36], another convex optimization problem which has many applications. Symmetric cones are studied in detail from the viewpoint of Riemannian geometry [18] and Hessian geometry

[51].

[Iteration-complexity bound based on an integral over the central trajectory (in particular for linear programming)]

Another thread of research stemmed from [28] concentrates on estimating the number of iterations of interior-point algorithms with a certain integral over the central trajectory, putting aside the Riemannian geometry. Many of these results are on primal-dual interior-point methods for linear programming [29, 30, 32, 34, 35, 57] which are known to be the best. The first concrete result is due to Sonnevend, Stoer and Zhao [53], who introduced a fundamental integral on the central trajectory which, roughly, is proportional to the iteration-complexity of the interior-point algorithms. Stoer and Zhao [74] subsequently established a rigorous bound of the number of iterations of a predictor-corrector type primal-dual algorithm based on the integral (see also [73]). Later Monteiro and Tsuchiya proved that the number of iterations is estimated "precisely" with the integral in an asymptotic sense [39].

On the other hand, in the middle of 90's, Vavasis and Ye developed an interesting interior-point method called layered-step interior-point method [71] (see also [33, 37] for variants). The algorithm is a modification of the Mizuno-Todd-Ye (MTY) predictor-corrector (PC) algorithm [34], where a step called the layered-step is taken once in a while instead of an ordinary predictor-step. A great feature of this algorithm is that its iteration-complexity is $O(n^{3.5}\log(\bar{\chi}_A + n))$, just depends on A but neither b nor c, where n is the number of nonnegative variables and $\bar{\chi}_A$ is a condition number of the coefficient matrix A of a linear program [16, 24, 54, 62, 63, 71], and bounded by $2^{O(L_A)}$, where L_A is the input bit length of A. Monteiro and Tsuchiya extensively studied relationship between the Vavasis-Ye algorithm and the standard MTY-PC algorithm, and succeeded to obtain an iteration-complexity bound of the MTY-PC algorithm which yet depends on b and c but in a weaker sense [38]. Finally, Monteiro and Tsuchiya established that the aforementioned integral introduced by Sonnevend, Stoer and Zhao is bounded by $O(n^{3.5}\log(\bar{\chi}^*_A + n))$ over the whole central trajectory [39], where $\bar{\chi}^*_A$ is a scaling invariant condition number of the coefficient matrix A introduced in [38] as an extension of $\bar{\chi}_A$ and $\bar{\chi}^*_A \le \bar{\chi}_A \le 2^{L_A}$ holds.

[Information geometry and interior-point algorithms]

Now we turn our attention on information geometry. Information geometry [3, 6, 41] is a differential geometrical framework successfully applied to many areas in informatics including statistics, signal processing, machine learning, statistical physics etc. Riemannian metric and two mutually dual connections are introduced based on a strongly convex function. In particular, the dually flat space which is "flat" with respect to the both connections plays fundamental role in information geometry.

One of the key observations in connecting interior-point algorithms to information geometry is that the trajectories appearing in the primal interior-point methods are "straight lines" under the gradient map of the logarithmic barrier function. This fact is observed by Bayer and Lagarias [8, 9], and Tanabe [56, 59]. Tanabe studied the map under the name of center-flattening transformation, in connection with differential geometrical framework for nonlinear programming (see, e.g., [55]). Iri [25] studied integrability of vector and multivector field associated with the interior-point methods generated by the logarithmic barrier function, which is somewhat connected to this structure.

The connection between information geometry and interior-point algorithms was firstly pointed out in Tanabe and Tsuchiya [60], in the context of linear programming. Specifically, they demonstrated that the mathematical structure of the interior-point algorithms for linear programs is very closely related to the dually flat space defined by the logarithmic barrier function. The affine-scaling trajectories, which contain the central trajectory as a special case, are dual geodesics. Tanabe further studied fiber structure of primal and dual problems in linear programming from information geometric viewpoint [58] (see also [59]).

Subsequently, Ohara introduced an information geometrical framework for symmetric positive semidefinite matrices defined by the logarithmic determinant function. See, for example, [48] for details. Ohara studied a class of directly solvable semidefinite programs from the viewpoint of information geometry and Euclidean Jordan algebra, and found that the concept of doubly autoparallel submanifold plays a fundamental role for such class [45, 46]. In these papers, he further developed a predictor-corrector type path-following algorithm for semidefinite programming based on information geometry, and analyzed a relation between the embedding curvature and performance of the predictor-step. The algorithm we will develop and analyze in this paper is based on this algorithm. Doubly autoparallel submanifold for symmetric cones is later studied in [47, 69].

Finally, we should mention that [17] introduces one parameter family of connections in the Riemannian geometrical setting defined by the self-concordant barrier, and studies behavior of the associated geodesics. This is exactly the same as α-connections in information geometry.

[**Other geometrical approaches**]

Curvature of the central trajectory for linear programs has been studied from various other viewpoints. Deza et al. [14] studied interesting behavior of the central trajectory for Klee-Minty cubes with numerous redundant constraints. Dedieu et al. [13] analyzed average total curvature of the central trajectory by using techniques from integral geometry and algebraic geometry. They analyze curvature of the central trajectory as a curve in (high)-dimensional Euclidean space, and this setting seems somewhat different from ours.

1.2 Outline and main results

Now we are ready to outline the structure and the main results of this paper. Sections 1 to 3 are preliminary sections to introduce information geometry, interior-point methods based on the Nesterov-Nemirovski self-concordant barriers.

The original results are presented from Section 4. We consider the information geometrical structure based on the p-logarithmically homogeneous self-concordant (i.e., p-normal) barrier function $\psi(x)$ associated with the cone Ω. The Riemannian metric is defined as the Hessian of ψ. The Legendre transformation $-\partial\psi(x)/\partial x$ defines a one-to-one mapping between Ω and Ω^*, and ∇- and ∇^*-connection are defined based on this structure. In particular, the interior $\mathcal{P}$ of the primal feasible region is a ∇-autoparallel submanifold, and the interior $\mathcal{D}$ of the dual feasible region is a ∇^*-autoparallel submanifold. The central trajectory γ is the path defined as the collection of the optimal solution $x(t)$ of the problem

$$\min\ t\langle c,\, x\rangle + \psi(x), \quad \text{s.t. } x \in (d+\mathbf{T})\cap \mathrm{cl}(\Omega), \tag{3}$$

where $t>0$ is a parameter. The interior-point algorithms follow this path approximately. It is known that $x(t)$ is an (p/t)-approximate solution, i.e, $\langle c, x(t)-x_{\rm opt}\rangle \le p/t$ ($x_{\rm opt}$ is an optimal solution). As one dimensional submanifold in Ω (or Ω^*), the central trajectory is written as $\gamma \equiv \{\gamma(t)|t\in(0,\infty]\}$, where $x(t)=x(\gamma(t))$. The central trajectory is characterized as $\gamma = \mathcal{P}\cap \mathrm{Hom}(\mathcal{D})$, the intersection of ∇-autoparallel submanifold $\mathcal{P}$ and ∇^*-autoparallel submanifold $\mathrm{Hom}(\mathcal{D})$, where $\mathrm{Hom}(\mathcal{D})$ is a "homogenization" of $\mathcal{D}$.

Then we develop a predictor-corrector path-following algorithm. We define the neighborhood $\mathcal{N}(\beta)\subset \mathrm{int}(\Omega)$ of the central trajectory γ, where $\beta\in(0,1)$ is a parameter to determine its width. We trace the path in the dual coordinate (∇^*-affine coordinate), i.e., in the space of the dual variables. One iteration of the algorithm consists of a predictor-step and a corrector-step. In the predictor-step, the iterate is supposed to be sufficiently close to the path γ and we take the largest step along the tangent of the path within the neighborhood $\mathcal{N}(\beta)$. Then the iterate is pulled back toward the central trajectory with the corrector step. This algorithm is developed based on information geometry, and a bit different from the traditional predictor-corrector algorithm in the literatures. We will prove polynomiality of the algorithm based on the theory of Nesterov and Nemirovski. Specifically, we will show that the algorithm is able to move from a neighbor point of $\gamma(t_1)$ to a neighbor point of $\gamma(t_2)$ ($t_1<t_2$) in $O(\sqrt{p}\log(t_2/t_1))$ iterations, going through the neighborhood $\mathcal{N}(\beta)$ of the path γ.

Then in Sections 5 and 6, we will develop the most interesting results in this paper. In Section 5, we will show that the number of iterations of the predictor-corrector algorithm developed in Section 4 with the neighborhood $\mathcal{N}(\beta)$ is written as follows, completely in terms of information geometry:

$$\begin{aligned}&(\#\text{ of iterations of the interior-point method to follow the central trajectory from } \gamma(t_1) \text{ to } \gamma(t_2))\\ &= \frac{1}{\sqrt{\beta}}\int_{t_1}^{t_2}\frac{1}{\sqrt{2}}\|H^*_{\mathcal{P}}(\dot\gamma,\dot\gamma)\|^{1/2}dt + \frac{o(1)}{\sqrt{\beta}}.\end{aligned}$$

Here, $H^*_{\mathcal{P}}(\cdot,\cdot)$ is the embedding curvature of the primal feasible region with respect the dual connection, and $\dot\gamma$ is the tangent vector of the central trajectory. We note that the above estimate becomes zero when $H^*_{\mathcal{P}}(\cdot,\cdot)=0$. This corresponds to the case where $\mathcal{P}$ is doubly autoparallel [45, 46, 69].

In Section 6, we apply the result to the concrete case of classical linear programming. With the help of the results developed by Monteiro and Tsuchiya [37, 38, 39] and Vavasis and Ye [71], we will show that the total curvature integral (improper integral from 0 to ∞) exists and is bounded as follows:

$$\int_0^\infty \frac{1}{\sqrt{2}}\|H^*_{\mathcal{P}}(\dot\gamma,\dot\gamma)\|^{1/2}dt = O(n^{3.5}\log(\bar\chi^*_A+n)),$$

where $\bar\chi^*_A$ is a scaling-invariant condition number of the coefficient matrix $A \in R^{n\times m}$ of the standard form linear program [38, 39]. Thus, surprisingly, the bound does not depend on b nor c. The condition number $\bar\chi^*_A$ is known to be $O(2^{L_A})$ where L_A is the input size of A, and for the class of combinatorial linear programs including network problems where A is a 0-1 matrix, we have the bound

$$\int_0^\infty \frac{1}{\sqrt{2}}\|H^*_{\mathcal{P}}(\dot\gamma,\dot\gamma)\|^{1/2}dt = O(n^{4.5}m),$$

just depends on the dimension of the problem. This result may have its own interest as a theorem in information geometry, describing a global structure of the central trajectory as a geometrical object. Through the analysis, we also provide a complete characterization of the primal-dual curvature studied in [39] in terms of information geometry. As a direct implication, we show that the primal (or) dual algorithm always require fewer iterations than primal-dual algorithms in view of the integral bound.

1.3 Note on Revision (March 2009)

We found that the embedding curvature was not expressed correctly in the first version released on December 2007 (the last two lines of p.10 and (10)). The expression was used in the proof of Lemma 5.3. We removed that part in revision and rewrote the proof of Lemma 5.3 (without using the expression). The lemma itself holds without any change of statement. Other typos were also corrected in revision.

2 Information Geometry and Dually Flat Space

In this section, we briefly describe a framework of information geometry. For details see [4, 6, 51].

[**Dually flat spaces**]

Let $\mathbf{E}$ and $\mathbf{E}^*$ be an n-dimensional vector space and its dual space, respectively. We denote by $\langle s, x\rangle$ the duality product of $x \in \mathbf{E}$ and $s \in \mathbf{E}^*$. Let $\{e_1,\dots,e_n\}$ be basis vectors of $\mathbf{E}$ and $\{e^1_*,\dots,e^n_*\}$ be its dual basis vectors of $\mathbf{E}^*$ satisfying $\langle e_i, e^j_*\rangle = \delta^j_i$. For $x \in \mathbf{E}$ we consider the affine coordinate system $(x^1,\dots,x^n)$ with respect to $\{e_1,\dots,e_n\}$, i.e, $x = \sum x^i e_i$. This coordinate system is referred to as x-coordinate. Similarly, for $s \in \mathbf{E}^*$ we consider the affine coordinate system $(s_1,\dots,s_n)$ with respect to $\{e^1_*,\dots,e^n_*\}$, i.e., $s = \sum s_i e^i_*$. This coordinate system is referred to as s-coordinate.

Let Ω be an open convex set in $\mathbf{E}$. We consider Ω as a manifold with global coordinate x as explained above and introduce dually flat structure on it. Let $\psi(x)$ be a smooth strongly convex function on Ω. Then the mapping $s(\cdot) : \Omega \to \mathbf{E}^*$ defined by:

$$s(\cdot) : x \mapsto s, \quad s_i = -\partial\psi/\partial x^i \tag{4}$$

is smoothly invertible on its image $\Omega^* \equiv s(\Omega) \subset \mathbf{E}^*$ because the Hessian matrix of ψ is positive definite. Thus, we can identify Ω^* with Ω, and call $(s_1, \ldots, s_n)$ *dual coordinates* of Ω. The set Ω^* becomes convex in $\mathbf{E}^*$ under appropriate regularity condition, e.g., if $\psi(x) \to \infty$ as $x \to \partial\Omega$. We illustrate the situation in Fig. 1.

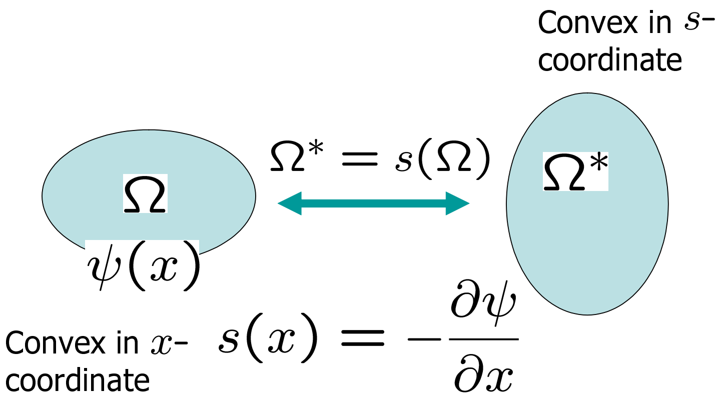

Figure 1: Dually flat space.

The conjugate function ψ^* on Ω^* is defined by

$$\psi^*(s) = \sup_{x \in \Omega}\{\langle -s, x\rangle - \psi(x)\}.$$

Note that the inverse mapping of $x(\cdot) : \Omega^* \to \Omega$ is represented by

$$x(\cdot) : s = (s_i) \mapsto x = (x^i), \quad x^i = -\partial\psi^*/\partial s_i. \tag{5}$$

In this paper we call the mapping $s(\cdot)$ and $x(\cdot)$ the *Legendre transformations.*

Riemannian metric tensor G on Ω defines an inner product of $T_x\Omega$, the tangent space of Ω at x. For basis vector fields $\partial/\partial x^i$ on Ω, their inner product are given by the components of G, i.e.,

$$(G_x)_{ij} = G\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right).$$

In information geometry, we define G by the Hessian matrix of ψ, i.e.,

$$(G_x)_{ij} \equiv \frac{\partial^2\psi}{\partial x^i \partial x^j}(x).$$

Then, for the basis vector fields $\partial/\partial s_i$, it holds that

$$(G_s)^{ij} \equiv G\left(\frac{\partial}{\partial s_i}, \frac{\partial}{\partial s_j}\right) = \frac{\partial^2 \psi^*}{\partial s_i \partial s_j}(s).$$

Note that we have $G_x^{-1} = G_s$ for $s = s(x)$, and further the Jacobi matrix for the coordinate transformation from x to s coincides with $-G_x$, i.e,

$$\frac{\partial s}{\partial x} \equiv \left(\frac{\partial s_i}{\partial x^j}\right) = -G_x.$$

For $V \in T_x\Omega$, the length $\sqrt{G_x(V,V)}$ of V is denoted by $\|V\|_x$.

Now we introduce affine connections on Ω, which determines structure of the manifold such as torsion and curvatures. The most significant feature of information geometry is that it invokes two affine connections ∇ and ∇^*, which accord with dualities in convex analysis, rather than the well-known Levi-Civita connection in Riemannian geometry. The affine connections ∇ and ∇^* are defined so that the x- and s-coordinates are, respectively, ∇- and ∇^*-*affine*, in other words, their associated covariant derivatives satisfy

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = 0, \quad \nabla^*_{\frac{\partial}{\partial s_i}} \frac{\partial}{\partial s_j} = 0, \quad \forall i, j. \tag{6}$$

Consequently, we also have the following alternative expressions in the other coordinates:

$$\nabla_{\frac{\partial}{\partial s_i}} \frac{\partial}{\partial s_j} = \sum_{k=1}^{n} \Gamma^{ij}_{k} \frac{\partial}{\partial s_k}, \quad \Gamma^{ij}_{k}(s) \equiv \sum_{k=1}^{n} G_{kl} \frac{\partial^3 \psi^*}{\partial s_i \partial s_j \partial s_l}.$$

$$\nabla^*_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} = \sum_{k=1}^{n} \Gamma^{*k}_{ij} \frac{\partial}{\partial x^k}, \quad \Gamma^{*k}_{ij}(x) \equiv \sum_{k=1}^{n} G^{kl} \frac{\partial^3 \psi}{\partial x^i \partial x^j \partial x^l}.$$

The definition (6) implies that both ∇ and ∇^* are flat connections, i.e., the associated torsions and Riemann-Christoffel curvatures vanish. Given the structure (G, ∇, ∇^*) constructed above, Ω is called a *dually flat* manifold. We say ψ and ψ^* *potential functions* of the structure.

[**Autoparallel Submanifolds and Embedding Curvature**]

Let $\mathcal{M}$ be the intersection of Ω and an arbitrary affine subspace $\mathcal{A} \subset \mathbf{E}$ represented in x-coordinate, i.e.,

$$\mathcal{M} \equiv \Omega \cap \mathcal{A}, \quad \mathcal{A} \equiv \{x | x = c_0 + \sum_{i=1}^{k} y^i c_i,\ c_i \in \mathbf{E}, y^i \in \mathbf{R}\}, \tag{7}$$

then, by the definition in (6), $\mathcal{M}$ is interpreted as a flat submanifold in terms of ∇. We call a submanifold $\mathcal{M}$ defined in (7) ∇-*autoparallel*. In particular, a one-dimensional ∇-autoparallel submanifold is called a ∇-*geodesic*. Similarly, a ∇^*-*autoparallel* submanifold $\mathcal{M}^*$ is defined by the intersection of Ω and $x(\mathcal{A}^*)$, where $\mathcal{A}^* \subset \mathbf{E}^*$ is an arbitrary affine subspace represented in s-coordinate, and we call a one-dimensional ∇^*-autoparallel submanifold a ∇^*-*geodesic*. Let $\mathcal{M}$ be a ∇-autoparallel submanifold of Ω and consider its homogenization in x-coordinate, i.e.,

$$\mathrm{Hom}(\mathcal{M}) \equiv \{x|x = tc_0 + \sum_{l=1}^{k} y^i c_i, t > 0\} = \bigcup_{t>0} t\mathcal{M}, \quad t\mathcal{M} \equiv \{x|x = tx', x' \in \mathcal{M}\}.$$

Then, $\mathrm{Hom}(\mathcal{M})$ is a ∇-autoparallel submanifold of Ω and $\{t, y^1, \ldots, y^k\}$ is a ∇-affine coordinates for $\mathrm{Hom}(\mathcal{M})$. An analogous notation is applied to a ∇^*-autoparallel submanifold in Ω using s-coordinate. We sometimes say x- and s-autoparallel meaning that ∇- and ∇^*-autoparallel.

Let $\mathcal{M}$ be a k-dimensional submanifold in the dually flat manifold Ω. We define the embedding curvature $H_{\mathcal{M}}(\cdot, \cdot)$ as follows. Since the tangent space $T_x\Omega$ at $x \in \mathcal{M}$ has the orthogonal decomposition with respect to the Riemannian metric G, i.e.,

$$T_x\Omega = T_x\mathcal{M} \oplus T_x\mathcal{M}^{\perp},$$

we can define the orthogonal projection $\Pi_x^{\perp} : T_x\Omega \to T_x\mathcal{M}^{\perp}$ at each x. For tangent vector fields X and Y on $\mathcal{M}$, let $H_{\mathcal{M}}(X, Y)$ be a normal vector field on $\mathcal{M}$ defined by

$$(H_{\mathcal{M}}(X, Y))_x = \Pi_x^{\perp}(\nabla_X Y)_x \in T_x\mathcal{M}^{\perp}.$$

at each x. Such a tensor field $H_{\mathcal{M}}$ is called the *(Euler-Schouten) embedding curvature* or the *second fundamental form* of $\mathcal{M}$ with respect to ∇. Similarly, we can introduce the dual embedding curvature $H^*_{\mathcal{M}}$ by replacing ∇ with ∇^*, i.e,

$$(H^*_{\mathcal{M}}(X, Y))_x = \Pi_x^{\perp}(\nabla^*_X Y)_x \in T_x\mathcal{M}^{\perp}.$$

It is shown that $\mathcal{M}$ is ∇-autoparallel (∇^*-autoparallel) *iff* $H_{\mathcal{M}} = 0$ ($H^*_{\mathcal{M}} = 0$).

For later use, we provide a concrete formula of $\Pi^{\perp}$ in x-coordinate and s-coordinate. Suppose that $T_x\mathcal{M} \subset \mathbf{E}$ is represented by the kernel of a certain linear operator $A : \mathbf{E} \to \mathbf{R}^m$, i.e., $T_x\mathcal{M} = \mathrm{Ker}A$. Define another operator $A^{\mathsf{T}} : \mathbf{R}^n \to \mathbf{E}^*$ by

$$\langle A^{\mathsf{T}} y, x\rangle = (y|Ax), \quad \forall y \in \mathbf{R}^m,$$

where $(\cdot|\cdot)$ is the standard inner product of $\mathbf{R}^m$. Then, the orthogonal projection $\Pi^{\perp} : \mathbf{E} \to (\mathrm{Ker}A)^{\perp} = T_x\mathcal{M}^{\perp}$ is

$$\Pi^{\perp} = G_x^{-1} A^{\mathsf{T}} (A G_x^{-1} A^{\mathsf{T}})^{-1} A = G_s A^{\mathsf{T}} (A G_s A^{\mathsf{T}})^{-1} A, \tag{8}$$

where $G_s = G_x^{-1}$, and G_x is regarded as an operator that maps $X = (X^i) \in \mathbf{E}$ to $S = (S_i) \in \mathbf{E}^*$:

$$G_x : X \mapsto S, \quad \text{where } S_i = \sum_j^n \frac{\partial^2 \psi}{\partial x^i \partial x^j} X^j.$$

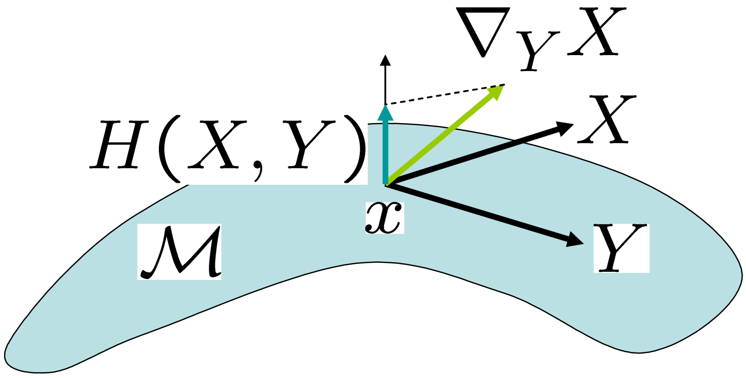

Figure 2: Embedding curvature.

The formula (8) is a coordinate expression of $\Pi^\perp$ in x-coordinate. $\Pi^\perp$ in s-coorinate is given by $G_x \Pi^\perp G_x^{-1}$, since $-G_x$ is the operator of coordinate transformation from x-coordinate to s-coordinate at the tangent space $T_x(\mathcal{M})$.

Remark: In the rest of the paper, we identify $\mathbf{E}$ and $T_x\Omega$, the tangent space at $x \in \Omega$, via the isomorphism: $e_i \mapsto (\partial/\partial x^i)_x$. Hence, for example, the sum $x + X$ of $x \in \Omega$ and $X \in T_x\Omega = \mathbf{E}$ makes sense in this identification, which would be often used to describe the algorithm later. Similarly, we also identify $\mathbf{E}^*$ and $T_s\Omega^*$, the tangent space at $s \in \Omega^*$.

[**Self-concordant functions**]

Finally we introduce widely known concept of the *self-concordant function* and *self-concordant barrier* [42], and describe their properties in terms of the above geometric notions. Let (G, ∇, ∇^*) be dually flat structure on Ω generated by the potential function ψ. If the inequality

$$\left| \langle X, \frac{\partial^3 \psi}{\partial x^3}(X, X) \rangle \right| \leq 2(G(X,X))^{3/2} = 2\|X\|_x^3,$$

or, equivalently,

$$\left| \sum_{i,j,k}^{n} \frac{\partial \psi^3(x)}{\partial x_i \partial x_j \partial x_k} X^i X^j X^k \right| \leq 2 \left(\sum_{i,j}^{n} \frac{\partial \psi^2(x)}{\partial x_i \partial x_j} X^i X^j \right)^{3/2}$$

holds over all $x \in \Omega$ and $X \in \mathbf{E}$, we call ψ a *self-concordant* function on Ω. The following property is a remarkable feature of the self-concordant function (see Appendix 1 of [42]):

$$\left| \langle X, \frac{\partial^3 \psi}{\partial x^3}(Y, Z) \rangle \right| \leq 2\|X\|_x \|Y\|_x \|Z\|_x. \quad \forall X,\ Y,\ Z \in \mathbf{E} \text{ and } x \in \Omega. \tag{9}$$

A self-concordant function ψ is said to be a *self-concordant barrier* if it satisfies

$$\psi(x) \to \infty \text{ as } x \to \partial\Omega.$$

If, in addition, the self-concordant barrier satisfies the condition

$$|\langle s(x), X\rangle| \le \sqrt{q}(G_x(X,X))^{1/2}, \quad \text{or, equivalently,} \quad \sum_i \left|\frac{\partial\psi(x)}{\partial x_i}X^i\right| \le \sqrt{q}\left(\sum_{i,j}^n \frac{\partial\psi^2(x)}{\partial x_i \partial x_j}X^i X^j\right)^{1/2}$$

for all $x \in \Omega$, then we call $\psi(x)$ q-self-concordant barrier on Ω.

Let Ω be an open convex cone. A barrier function $\psi(x)$ on Ω is called a q-logarithmically homogeneous barrier if

$$\psi(tx) = \psi(x) - q\log t$$

holds for $t > 0$. If $\psi(x)$ is a q-logarithmically homogeneous barrier, then the Legendre transformation $s(\cdot)$ defined by (4) satisfies the following equalities for all $x \in \Omega$, $X \in \mathbf{E}$ and $t > 0$:

$$s(tx) = t^{-1}s(x), \tag{10}$$

$$G_{tx} = t^{-2}G_x,$$

$$\langle s(x), x\rangle = q, \tag{11}$$

$$\langle s(x), X\rangle = G(x, X).$$

See Proposition 2.3.4 of [42] for their proofs. A self-concordant barrier is referred to as a *q-normal barrier* if it is q-logarithmically homogeneous. A q-normal barrier is known to be a q-self-concordant barrier. A remarkable feature of the q-normal barrier is that the conjugate function $\psi^*(s)$ again becomes a q-normal barrier with domain Ω^* and Ω^* becomes the dual cone of Ω. The Legendre transformation $x(\cdot)$ defined by (5) satisfies the analogous relations for all $s \in \Omega^*$, $S \in \mathbf{E}^*$ and $t > 0$:

$$x(ts) = t^{-1}x(s), \tag{12}$$

$$G_{ts} = t^{-2}G_s, \tag{13}$$

$$\langle x(s), s\rangle = q, \tag{14}$$

$$\langle x(s), S\rangle = G(s, S). \tag{15}$$

3 Self-concordant Barrier Function, the Newton method, and Interior-point Algorithms

In this section, we describe basic results on the polynomial-time interior-point algorithms for general convex programs by Nesterov and Nemirovski [42]. We also explain the algorithm in words of information geometry. For later convenience, we will use a bit different notations from the previous sections.

Let $\mathbf{E}$ be a k-dimensional vector space, and let $\mathcal{C}$ be open convex set of $\mathbf{E}$, and let y be an affine coordinate of $\mathbf{E}$. Let ϕ be a self-concordant barrier whose domain is $\mathcal{C}$. Information geometric structure explained in the previous section is introduced by taking ϕ as the potential function.

The Riemannian metric is denoted by G, and the dual coordinate is introduced by the Legendre transformation $g = -\partial\phi/\partial y$.

For $y \in \mathcal{C}$ and $r \geq 0$, we define an ellipsoid

$$D(y, r) \equiv \{v = y + V \mid \|V\|_y \leq r,\ V \in \mathbf{E}\}.$$

This ellipsoid is referred to as the *Dikin ellipsoid* (with radius r).

It is known that the local norm $\|\cdot\|_y$ changes moderately and is compatible among each other within the Dikin ellipsoid of radius 1.

Theorem 3.1 *(Theorem 2.1.1 of [42], see also [50] and [26]) Let $u \in \mathcal{C}$. If $v \in D(u, r)$, $r < 1$, then, $v \in \mathcal{C}$, and, for any $V \in \boldsymbol{E}$, we have*

$$(1 - r)\|V\|_u \leq \|V\|_v \leq \frac{1}{1 - r}\|V\|_u.$$

Let $\hat{g} \in g(\mathcal{C}) = \{g(y) \mid y \in \mathcal{C}\}$, and we consider computing a point $\hat{y} \in \mathcal{C}$ such that the Legendre transformation $g(\hat{y}) = \hat{g}$. This is the problem of computing the inverse Legendre transformation. Generally there is no explicit formula for the inverse Legendre transformation. Such point is characterized as the optimal solution to the optimization problem

$$\min_y\ \langle \hat{g},\ y\rangle + \phi(y), \quad \text{s.t. } y \in \mathcal{C}, \tag{16}$$

and we compute it with the Newton method. The minimum point $y \in \mathcal{C}$ of ϕ satisfying the condition $\hat{g} = g(y) = 0$ is referred to as the *analytic center*, and plays an important role in the theory of interior-point algorithms (Fig. 3).

The Newton displacement vector $N_y(\hat{g})$ at y of the Newton method to solve (16) is given as the unique vector satisfying

$$\langle \hat{g}, V\rangle = \langle g(y), V\rangle - G_y(N_y(\hat{g}), V) \quad \forall V \in \mathbf{E}.$$

The norm $\|N_y(\hat{g})\|_y$ of the Newton displacement vector measured with the Riemannian metric G at y is referred to as the *Newton decrement*.

We have the following fundamental result which guarantees that the inverse Legendre transformation is computed efficiently with the Newton method.

Theorem 3.2 *(Theorem 2.2.2 of [42], see also [50] and [26]) Suppose that $\|N_y(\hat{g})\|_y = \beta < 1$, and let y^+ be the point obtained by performing one step of the Newton iteration, i.e.,*

$$y^+ = y + N_y(\hat{g}).$$

Then

$$\|N_{y^+}(\hat{g})\|_{y^+} \leq \frac{\beta^2}{(1 - \beta)^2},$$

and therefore, if $\beta \leq 1/4$,

$$\|N_{y^+}(\hat{g})\|_{y^+} \leq \frac{16\beta^2}{9}.$$

The sequence $\{y^k\}$ generated by the Newton method initiated at a point y^0 such that $\|N_{y^0}(\hat{g})\|_{y^0} < 1$ converges to the point $y^\infty \in \mathcal{C}$ such that $g(y^\infty) = \hat{g}$.

It is easy to see that $N_y(0)$ is the Newton displacement vector for minimizing $\phi(y)$ to obtain the analytic center. We describe a few properties $N_y(0)$ which is utilized later.

Proposition 3.3 *If ϕ is a q-self-concordant barrier function, then, for any $y \in \mathcal{C}$, we have*

$$\|N_y(0)\|_y \leq \sqrt{q}. \tag{17}$$

This fact is readily seen from the definition of q-selfconcordant barrier function and the Newton displacement vector. Here we note that $N_y(0)$ is well-defined even when the feasible region is not bounded hence the analytic center does not exist. Even in such a case, the above proposition holds. The following property about $N_y(0)$ is used later. Let $y \in \mathcal{C}$, $g_0 = g(y)$, $\alpha \in \mathbf{R}$. Then, we have

$$N_y(\alpha g_0) = (1-\alpha)N_y(0), \quad \|N_y(\alpha g_0)\|_y = |\alpha - 1|\|N_y(0)\|. \tag{18}$$

This readily follows from the general property of the Newton displacement vector.

Based on Theorems 3.1 and 3.2, we can prove the following proposition which is used later. See Appendix for the proof.

Proposition 3.4 *Let $y_1, y_2 \in \mathcal{C}$, and let $g_1 = g(y_1), g_2 = g(y_2)$. Let*

$$\eta = \|N_{y_1}(g_2)\|_{y_1}, \quad \zeta = \|y_2 - y_1\|_{y_1}.$$

Suppose that $\eta \leq 1/16$. Then, the following holds.

1. *We have*

$$1 - 8\eta \leq \frac{\zeta}{\eta} \leq 1 + 8\eta. \tag{19}$$

2. *In addition, if*

$$\zeta \leq \frac{1}{64},$$

then,

$$1 - 22\zeta \leq \frac{\eta}{\zeta} \leq 1 + 22\zeta.$$

Let $b \in \mathbf{E}^*$ be a linear functional of $\mathbf{E}$, and let us consider the following minimization problem:

$$\min \langle b, y \rangle \quad \text{s.t. } y \in \text{cl}(\mathcal{C}). \tag{20}$$

Based on the self-concordant barrier function ϕ with the barrier parameter q, the path-following interior-point algorithm solves the problem by following the *central trajectory* [23, 31, 32, 49, 52] defined as the set of optimal solution $y(t)$ of the following problem with parameter $t(> 0)$ (Fig. 3):

$$\min\ t\langle b, y\rangle + \phi(y), \quad \text{s.t. } y \in \text{cl}(\mathcal{C}).$$

The point $y(t)$ satisfies the equation

$$tb + \frac{\partial \phi}{\partial y} = 0, \text{ or equivalently, } tb - g(y) = 0.$$

Let y_{opt} be an optimal solution of (20). It is known that

$$\langle b, y(t)\rangle - \langle b, y_{\text{opt}}\rangle \leq \frac{\sqrt{q}}{t}.$$

The path-following algorithm is able to follow the central trajectory from the neighbor of $y(t_1)$ to that of $y(t_2)$, where $t_2 > t_1$, in $O(\sqrt{q}\log(t_2/t_1))$-iterations. Later we will address a version of such an interior-point path-following algorithm.

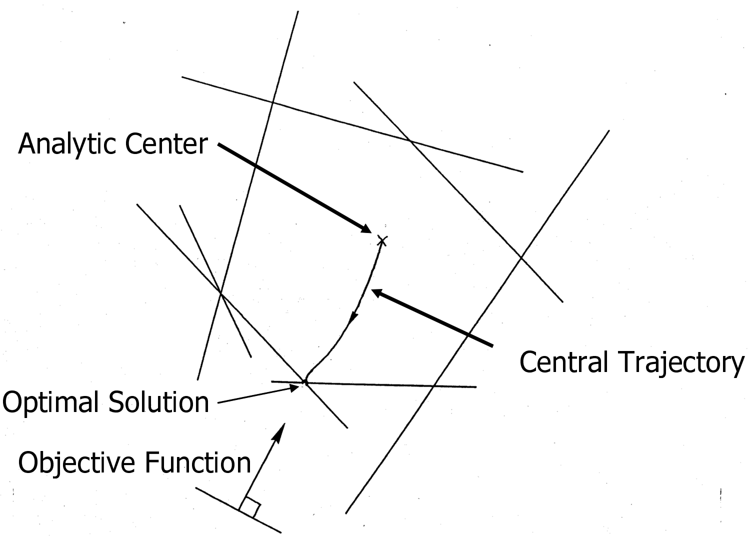

Figure 3: Central trajectory and analytic center.

Now we introduce two vector fields which play important roles in the theory of interior-point algorithms, namely, V_y^{aff} and V_y^{cen}. The vector field V_y^{aff} is defined as the negative gradient of the objective function b with respect to the metric G and is referred to as *the affine-scaling vector field* [15, 16]. The vector field V_y^{aff} is characterized as the one satisfying the following condition at each $y \in \mathcal{C}$:

$$G_y(V_y^{\text{aff}}, V) = -\langle b, V\rangle, \quad \forall V \in \mathbf{E}.$$

We note that the vector field $V_y^{\rm aff}$ is nothing but the *natural gradient* [5] known in neural networks and machine learning. The interior-point algorithm based on the affine-scaling vector field is called the affine-scaling algorithm [7, 15, 16, 64, 65, 67, 70]. The affine-scaling algorithm is probably one of the earliest instances where the idea of natural gradient is proved to be important in designing an efficient algorithm.

The other vector field $V_y^{\rm cen}$ is the Newton displacement vector field $N_y(0)$ for minimizing the barrier function ϕ. The vector field $V_y^{\rm cen}$ is referred to as *the centering vector field.*

The affine-scaling vector flow and the centering vector flow are shown in Fig. 4. It is easy to see that the integral curves of $V_y^{\rm aff}$ and $V_y^{\rm cen}$ are ∇^*-geodesics.

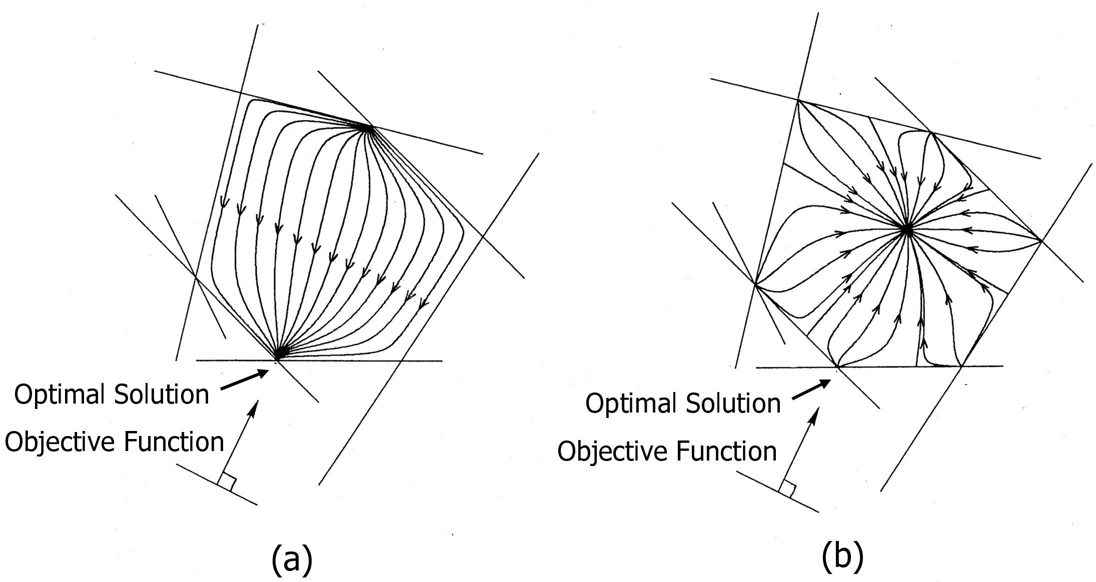

Figure 4: Integral curves of the affine-scaling vector field and the centering vector field for linear programming (they are ∇^*-geodesics). (a) the affine scaling vector integral curves; centering vector integral curves.

The central trajectory γ is characterized exactly as the integral curve of the affine-scaling vector field starting from the analytic center which also is an integral curve of the reversed centering vector field. On $y(t) \in \gamma$, we have the relation

$$\dot{y}(t) = V_y^{\rm aff} = -\rho(y) V_y^{\rm cen}, \quad \rho(y) > 0.$$

This means that $V_y^{\rm aff}$ and/or $-V_y^{\rm cen}$ are good approximations to the tangent $\dot{y}(t)$ of the central trajectory in a sufficiently small neighborhood of the central trajectory. This has the following meaning in constructing an implementable path-following algorithm.

If an iterate y is exactly on γ, $y = y(t)$, say, and if we develop a predictor-corrector type algorithm to trace γ, then definitely the direction which should be adapted for predictor-step is the tangent $\dot{y}(t)$. However, if we trace γ numerically by an iterative procedure, iterates cannot be exactly on the central trajectory. Generically, we define a neighborhood of the central trajectory, and iterates are generated in this neighborhood. In such a case, we need an appropriate substitute for $\dot{y}(t)$. The vector field V_y^{aff} and $-V_y^{\mathrm{cen}}$ may well be used for such purpose.

Finally, we consider the following situation for later consideration. Let $\mathbf{E}$ be an n-dimensional vector space, and $\Omega \subset \mathbf{E}$ be the open convex cone equipped with the q-self-concordant barrier function ψ, and consider the information geometric structure (G, ∇, ∇^*) based on ψ (We take ψ as the potential). Let $a^0 \in \mathbf{E}^*$, and let $a_i,\ i = 1, \ldots, k$ be linearly independent elements of $\mathbf{E}^*$, and let $A : \mathbf{E} \to \mathbf{R}^k$ as

$$Ax = \begin{pmatrix} \langle a_1, x\rangle \\ \vdots \\ \langle a_k, x\rangle \end{pmatrix}. \tag{21}$$

Let

$$\mathcal{S} = \{s = s(y) \mid s(y) = a_0 - \sum_{i=1}^{k} a_i y^i,\ y \in \mathbf{R}^k\} \cap \Omega^* = (a_0 + \mathbf{S}^*) \cap \Omega^*,$$

where $\mathbf{S}^* = \{V \in \mathbf{E}^* \mid V = -\sum_{i=1}^k a_i y^i,\ y \in \mathbf{R}^k\}$. Then $\mathcal{S}$ becomes a k-dimensional ∇^*-autoparallel submanifold and $\phi(y) = \psi^*(a_0 - \sum_{i=1}^k a_i y^i)$ becomes the q-self-concordant barrier function whose domain is

$$\mathcal{C} = \{y \in \mathbf{R}^k |\ s(y)\ \in \Omega^*\}.$$

We may consider information geometric structure on $\mathcal{C}$ induced by the potential function $\phi(y)$. The induced inner product is written as, for two tangent vectors V_y and W_y, $G_{s(y)}(A^\top V_y, A^\top W_y)$, where $A^\top$ is the mapping from $T_y\mathcal{C}$ to $T_{s(y)}\mathcal{S} \subset T_{s(y)}\Omega$ defined as

$$V_{s(y)} \equiv A^\top V_y = \sum_i a^i V_{y_i}.$$

We define $\|V_y\|_y$ as $G(A^\top V_y, A^\top V_y)$. Then,

$$\|V_{s(y)}\|_{s(y)} = \|A^\top V_y\|_{s(y)} = \|V_y\|_y.$$

Let $h \in \mathbf{E}$, and consider the convex optimization problem

$$\min\ \langle h,\ s\rangle + \psi^*(s), \quad s \in \mathcal{S}(= a^0 + S^*). \tag{22}$$

This is essentially equivalent to the following convex optimization problem over $\mathcal{C}$:

$$\min\ -\langle f,\ y\rangle + \phi(y), \quad \text{s.t. } y \in \mathcal{C}, \tag{23}$$

where $f = Ah$. The Newton decrement $\|N_y(f)\|_y$ is $\|A^\top N_y(f)\|_{s(y)}$.

On the other hand, the Newton displacement vector $N_s^{\mathbf{S}^*}(h)$ for (22) at $s \in \mathcal{S}$ is directly defined as the vector minimizing the following second-order approximation of the objective function.

$$\min \ \langle h, V\rangle + \frac{1}{2}G(V,V), \ V \in \mathbf{S}^*,$$

where G is the Hessian of ψ^* at $s \in \Omega$ and plays the role of metric. Then, we have the following proposition. The proof is omitted.

Proposition 3.5 *Under the setting above, let $s = s(y)$. Then,*

$$N_s^{\mathbf{S}^*}(h) = -A^\mathsf{T} N_y(-f), \quad \|N_s^{\mathbf{S}^*}(h)\|_s = \|N_y(f)\|_y.$$

The following is a corollary of Theorem 3.3.

Corollary 3.6 *Suppose that $\|N_s(h)\|_s \le \beta < 1$, and let s^+ be the point obtained by performing one step of the Newton iteration, i.e.,*

$$s^+ = s + N_s^{\mathbf{S}^*}(h).$$

Then

$$\|N_{s^+}^{\mathbf{S}^*}(h)\|_{s^+} \le \frac{\beta^2}{(1-\beta)^2},$$

and therefore, if $\beta \le 1/4$,

$$\|N_{s^+}^{\mathbf{S}^*}(h)\|_{s^+} \le \frac{16\beta^2}{9},$$

The sequence $\{s^k\}$ generated by the Newton method such that $\|N_{s^0}^{\mathbf{S}^*}(h)\|_{s^0} < 1$ converges to the point $s^\infty \in \mathcal{S}$, which is the unique optimal solution to (22).

Before proceeding to the next session, we derive the Newton displacement vector $N_y(-f)$ and $N_s^{\mathbf{S}^*}(h)$, and the Newton decrement $\|N_y(f)\|_y = \|N_s^{\mathbf{S}^*}(h)\|_s$. Taking account that $\partial^2\psi^*(s)/\partial s\partial s = G_s = G_x^{-1}$ and $\partial^2\phi(y)/\partial y\partial y = AG_sA^\mathsf{T}$, we see that the Newton displacement vector for the optimal solution of (23) is written as

$$N_y(-f) = (AG_sA^\mathsf{T})^{-1}(f + g(y)) = (AG_sA^\mathsf{T})^{-1}(f + A\frac{\partial\psi^*(s)}{\partial s}) = (AG_sA^\mathsf{T})^{-1}(f - Ax(s)).$$

The Newton displacement vector for (22) is derived as $N_s^{\mathbf{S}^*}(d) = -A^\mathsf{T} N_y(-f)$, and the Newton decrement becomes

$$\|N_s^{\mathbf{S}^*}(d)\|_s \equiv \sqrt{(f - Ax(s))^\mathsf{T}(AG_sA^\mathsf{T})^{-1}(f - Ax(s))}. \tag{24}$$

The following proposition is an analogue of Proposition 3.4.

Proposition 3.7 *Let $s_1, s_2 \in \mathcal{D}$, and let s_2 be the optimal solution of (22). Let*

$$\eta = \|N_{s_1}^{\mathbf{S}^*}(d)\|_{s_1}, \quad \zeta = \|s_2 - s_1\|_{s_1}.$$

Suppose that $\eta \le 1/16$. Then, the following holds.

1. We have

$$1 - 8\eta \leq \frac{\zeta}{\eta} \leq 1 + 8\eta.$$

2. In addition, if

$$\zeta \leq \frac{1}{64},$$

then,

$$1 - 22\zeta \leq \frac{\eta}{\zeta} \leq 1 + 22\zeta.$$

Proof. The proof goes similar to Proposition 3.4, just by using Corollary 3.6. ■

4 Main Idea and a Path-following Algorithm

In this section, we present a framework of treating conic linear programs based on information geometry. Then we develop a predictor-corrector type path-following algorithm, and state a polynomiality result of the algorithm (proof is given in the appendix).

4.1 Framework

Now we return to the dual pair of the problems (1) and (2). Let

$$\mathcal{P} \equiv (d + \mathbf{T}) \cap \Omega \text{ and } \mathcal{D} \equiv (c + \mathbf{T}^*) \cap \Omega^*.$$

Then (1) and (2) are written as

$$\min\ \langle c,\, x\rangle, \quad \text{s.t. } x \in \mathrm{cl}(\mathcal{P}), \tag{25}$$

and

$$\min\ \langle d,\, s\rangle, \quad \text{s.t. } s \in \mathrm{cl}(\mathcal{D}), \tag{26}$$

respectively. It is easy to see that, for any $x \in \mathrm{cl}(\mathcal{P})$ and $s \in \mathrm{cl}(\mathcal{D})$, we have $\langle x, s\rangle \geq 0$.

We also assume that

$$\mathcal{P} \neq \emptyset, \quad \mathcal{D} \neq \emptyset. \tag{27}$$

This is a standard assumption, and under this assumption, (25) and (26) have optimal solutions satisfying the following condition:

$$\langle x,\, s\rangle = 0,\ x \in \mathrm{cl}(\mathcal{P}),\ s \in \mathrm{cl}(\mathcal{D}).$$

Let $\psi(x)$ be a p-normal barrier whose domain is Ω. Then the conjugate function $\psi^*(s)$ of $\psi(x)$ is a p-normal barrier whose domain is Ω^* [42]. Based on $\psi(x)$ and $\psi^*(s)$, we introduce the central trajectories of (25) and (26).

Associated with (25), we consider the following optimization problem with parameter t

$$\min\ t\langle c,\, x\rangle + \psi(x) \ \text{ s.t. } x \in \mathrm{cl}(\mathcal{P}). \tag{28}$$

The optimality condition of this problem is written as:

$$\left(tc + \frac{\partial \psi}{\partial x} =\right) tc - s(x) \in \mathbf{T}^*, \quad x \in d + \mathbf{T}, \quad x \in \Omega. \tag{29}$$

Let $x_{\mathcal{P}}(t)$ be the unique optimal solution to (28), and $\gamma_{\mathcal{P}}(t)$ be the point in Ω expressed as $x_{\mathcal{P}}(t) = x(\gamma_{\mathcal{P}}(t))$ in x-coordinate. We define the central trajectory $\gamma_{\mathcal{P}}$ as:

$$\gamma_{\mathcal{P}} = \{\gamma_{\mathcal{P}}(t) |\ t \in [0, \infty)\}.$$

$\gamma_{\mathcal{P}}(t)$ converges to the optimal solution of (25) as $t \to \infty$.

Similarly, we consider the following optimization problem with parameter t associated with (26).

$$\min\ t\langle d,\, s\rangle + \psi^*(s)\ \text{ s.t. } s \in \mathrm{cl}(\mathcal{D}). \tag{30}$$

The optimality condition for this problem is

$$\left(td + \frac{\partial \psi}{\partial s} =\right) td - x(s) \in \mathbf{T}, \quad s \in c + \mathbf{T}^*, \quad s \in \Omega^*. \tag{31}$$

Let $s_{\mathcal{D}}(t)$ be the unique optimal solution to (28), and $\gamma_{\mathcal{D}}(t)$ be the point in Ω expressed as $s_{\mathcal{D}}(t) = s((\gamma_{\mathcal{D}}(t))$ in s-coordinate. We define the central trajectory $\gamma_{\mathcal{D}}$ as:

$$\gamma_{\mathcal{D}} = \{\gamma_{\mathcal{D}}(t) |\ t \in [0, \infty)\}.$$

$\gamma_{\mathcal{D}}(t)$ converges to the optimal solution of (26) as $t \to \infty$.

We also note that $\gamma_{\mathcal{P}}(t)$ (in s-coordinate) and $\gamma_{\mathcal{D}}(t)$ (in x-coordinate) are alternatively characterized as the optimal solutions of the following optimization problems.

$$\min\ \langle d,\, s\rangle + \psi^*(s)\ \text{ s.t. } s \in \mathrm{cl}(t\mathcal{D}). \tag{32}$$

$$\min\ \langle c,\, x\rangle + \psi(x)\ \text{ s.t. } x \in \mathrm{cl}(t\mathcal{P}).$$

These relations are readily verified by writing down the optimality conditions and using the definition of the Legendre transformation. Interior-point algorithms approach the optimal solutions by tracing the central trajectories $\gamma_{\mathcal{P}}$ and $\gamma_{\mathcal{D}}$.

Now we consider the information geometric structure on Ω as a dually flat manifold taking the p-normal barrier function $\psi(x)$ as the potential function. We identify Ω and Ω^* by the Legendre transformation. Then it is easy to see that $\mathcal{P}$ is an $(n-m)$-dimensional ∇-autoparallel (x-autoparallel) submanifold and $\mathcal{D}$ is an m dimensional ∇^*-autoparallel (s-autoparallel) submanifold. Note that a ∇-autoparallel (∇^*-autoparallel) submanifold is not necessarily ∇^*-autoparallel (∇-autoparallel). Thus, in view of x-coordinate, the primal feasible region $\mathcal{P}$ is an autoparallel manifold and the dual feasible region $\mathcal{D}$ is a curved submanifold, while, in view of s-coordinate, $\mathcal{D}$ is an autoparallel manifold and $\mathcal{P}$ is a curved submanifold. See Fig. 5.

The intersection of the two submanifolds is a unique point which is a point on the central trajectory. We have the following proposition. We omit the proof here.

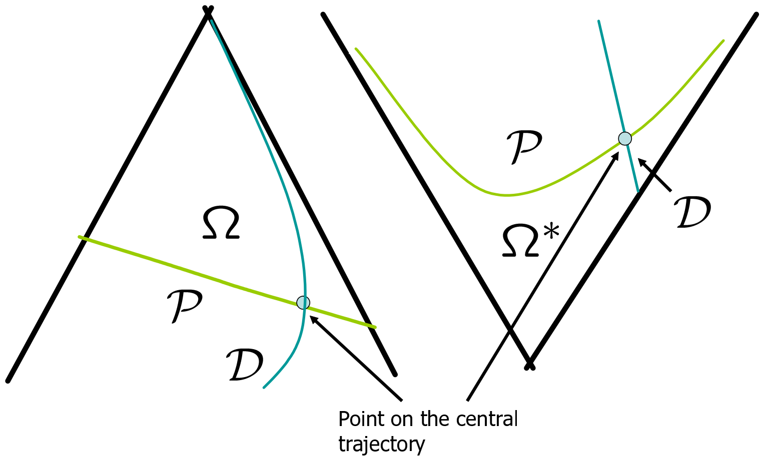

Figure 5: Primal and dual feasible regions in x-coordinate (left) and s-coordinate (right) .

Proposition 4.1 *$\mathcal{P} \cap \mathcal{D}$ is nonempty and is a unique point iff the condition (27) is satisfied, and written as the optimal solution of (28) with $t = 1$ in x-coordinate, and written as the optimal solution of (30) with $t = 1$ in s-coordinate (see Fig. 5).*

In view of this proposition, we have the following characterization of $\gamma_{\mathcal{P}}$ and $\gamma_{\mathcal{D}}$ as the intersection of two submanifolds:

$$\gamma_{\mathcal{P}}(t) = \mathcal{P} \cap t\mathcal{D} \text{ and } \gamma_{\mathcal{P}} = \mathcal{P} \cap \mathrm{Hom}(\mathcal{D}),$$

and

$$\gamma_{\mathcal{D}}(t) = \mathcal{D} \cap t\mathcal{P} \text{ and } \gamma_{\mathcal{D}} = \mathcal{D} \cap \mathrm{Hom}(\mathcal{P}).$$

The situation is illustrated in Fig. 6 for the primal central trajectory $\gamma_{\mathcal{P}}$.

It is well-known that the following result holds:

$$\langle x_{\mathcal{P}}(t),\ s_{\mathcal{D}}(t)\rangle = \frac{p}{t}, \tag{33}$$

and

$$x_{\mathcal{P}}(t) = \frac{x(s_{\mathcal{D}}(t))}{t} \text{ and } s_{\mathcal{D}}(t) = \frac{s(x_{\mathcal{P}}(t))}{t}. \tag{34}$$

(The relation (34) holds from (29), (31), (10) and (12). The relation (33) holds from (34), (11) and (14)).

In the following, we mostly deal with the central trajectory $\gamma_{\mathcal{P}}$ for the primal problem. We introduce a path-following algorithm to trace $\gamma_{\mathcal{P}}$ in the dual space $\Omega^* \subset \mathbf{E}^*$. The primal feasible region $\mathcal{P}$ is a curved submanifold embedded in the dual space Ω^*, and we follow the path $\mathcal{P} \cap \mathrm{Hom}(\mathcal{D})$ in Ω^*, utilizing s-coordinate.

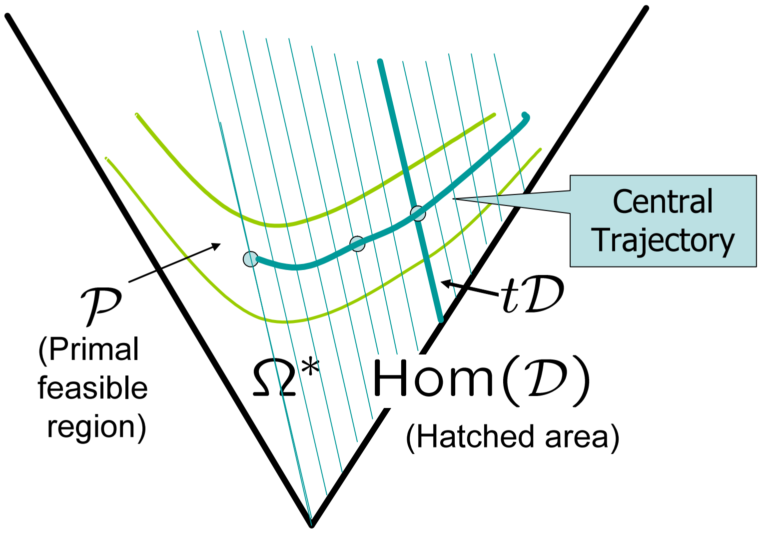

Figure 6: Central trajectory.

Let $a_1, \ldots, a_m \in \mathbf{E}^*$ be the basis of $\mathbf{T}^*$, and we define the linear operator from $\mathbf{E} \to \mathbf{R}^m$ as in (21). Then $\mathcal{D} = c + \mathbf{T}^*$ is written as, in s-coordinate,

$$\mathcal{D} = \{s \in \Omega^* \;|s = c - A^{\mathsf{T}}y, \quad y \in \mathbf{R}^m\}.$$

On the other hand, for $b \in \mathbf{R}^m$ satisfying $Ad = b$ we can express $\mathcal{P}$ as, in x-coordinate,

$$\mathcal{P} = \{x \in \Omega \;|Ax = b\}.$$

In the following, we use x-coordinate and s-coordinate as coordinate systems to carry out concrete calculations. Furthermore, we will use vector and matrix notations rather than traditional tensor notations with lower and upper indices. We abuse notations of identifying vectors and linear operators with its coordinate representations.

Now we derive the differential equation of the central trajectory $\gamma_{\mathcal{P}}$ written in the s-coordinate. The point $\gamma_{\mathcal{P}}(t)$ on the central trajectory $\gamma_{\mathcal{P}}$ in x-coordinate is the optimal solution of

$$\min \; t\langle c, \; x\rangle + \psi(x) \text{ s.t. } Ax = b.$$

The optimality condition implies that

$$tc - s = A^{\mathsf{T}}y, \; Ax = b, \; s = -\frac{\partial \psi(x)}{\partial x}.$$

One more differentiation with respect to t yields that

$$c + G_x\dot{x} = A^{\mathsf{T}}\dot{y}, \quad A\dot{x} = 0,$$

where $G_x = -\partial s/\partial x = \partial^2\psi/\partial x\partial x$. Note that G_x is the metric with respect to x-coordinate. Multiplying the first equality by G_x^{-1} and then multiply A from left we have

$$\dot{y} = (AG_x^{-1}A^\top)^{-1}AG_x^{-1}c \tag{35}$$

and

$$\dot{s} = -G_x\dot{x} = (G_x - A^\top(AG_x^{-1}A^\top)^{-1}A)G_x^{-1}c = G_x(I-\Pi^\perp)G_x^{-1}c = G_s^{-1}(I-\Pi^\perp)G_s c. \tag{36}$$

Let us consider the linear approximation

$$s_L(t') = s(t) + (t'-t)\dot{s}(t)$$

to the trajectory $\gamma_{\mathcal{P}}$ at $s(t) \in t\mathcal{D}$. It readily follows from (35) and (36) that

$$s_L(t') \in t'\mathcal{D} \text{ (as long as } s_L(t') \in \Omega^*). \tag{37}$$

Here we make an important observation. Though we derived it as the tangent of the central trajectory, the right hand side of (36) is well-defined for any $x \in \Omega$. Therefore, we consider the vector field V^{ct} which is written, in s-coordinate, as the righthand side of (36). This vector field is well-defined over Ω. The integral curves of this vector field describe all central trajectories of the problems of the form

$$\min\ \langle c, x\rangle,\quad \text{s.t. } x \in (d' + \mathbf{T}) \cap \mathrm{cl}(\Omega), \tag{38}$$

including (1). We have the following proposition. The proof is omitted.

Proposition 4.2 *The following holds:*

1. *Let γ be the maximal integral curve of V^{ct} passing through $d' \in \Omega$. Then, $\gamma \cap \mathrm{Hom}(\mathcal{D})$ is the central trajectory of the linear program (38).*

2. *Suppose that (38) has an interior feasible solution, i.e., a feasible solution x such that $x \in \Omega$. Then, its central trajectory is written as an integral curve of V^{ct}.*

Finally, we make an observation about the second derivative of the integral curve of V^{ct}. Let $s(t)$ be an integral curve of V_s^{ct}. Then $\dot{s}$ is written as in (36). The second derivative is written as follows as covariant derivative of V^{ct}.

$$\ddot{s}(t) = \nabla^*_{V_s^{\mathrm{ct}}} V_s^{\mathrm{ct}}. \tag{39}$$

This readily follows from that s-coordinate is ∇^*-affine. In this respect, the second derivative of the curves form a vector field. We denote this vector field as $\ddot{s}(\cdot)$.

4.2 Path-following Algorithm

In this section, we introduce the path-following algorithm we will analyze. The algorithm follows the central trajectory $\gamma_{\mathcal{P}}$ with a homotopy Newton method. We denote by $\mathcal{N}(\beta)$ the neighborhood of the central trajectory $\gamma_{\mathcal{P}}$ where $\beta \in (0, 1)$ is a scalar parameter to determine its width. The iterates are generated in $\mathcal{N}(\beta)$. Smaller β yields smaller neighborhood. A more precise definition of $\mathcal{N}(\beta)$ is given later.

[Ideal algorithm] (See Fig. 7)

1. Let $\gamma_{\mathcal{P}}(t)$ be a point on the central trajectory $\gamma_{\mathcal{P}}$ with parameter t as defied in (36). Let $s(t) = s(\gamma_{\mathcal{P}}(t))$, and let $\dot{s}(t)$ be the tangent direction at t as defined in (36).

2. (Predictor step)

   Let $s_L(t+\Delta t) = s(t)+\Delta t\dot{s}$, and let $\Delta t_{\max} > 0$ be the maximum Δt such that $s_L(t+\Delta t) \in \mathcal{N}(\beta)$. ($s_L(t + \Delta t) \in (t + \Delta t)\mathcal{D}$ as we observed in (37).)

3. (Corrector step) Move to $s(t + \Delta t_{\max})$ from $s_L(t + \Delta t_{\max})$.

4. $t := t + \Delta t$ and return to step 1.

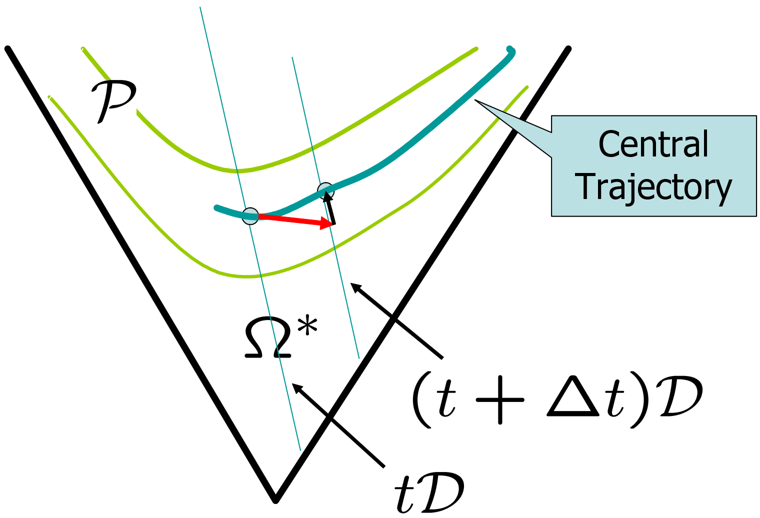

Figure 7: Path-following algorithm.

In this ideal algorithm, we made an unrealistic assumption that the predictor-step is performed from a point exactly on the central path. In reality, we cannot do this since we cannot perform perfect centering. In order to construct an implementable algorithm, we need to resolve this problem. We also need to define the neighborhood of the central trajectory. Now we explain how to deal with these issues.

[Predictor-step]

The central trajectory $\gamma_{\mathcal{P}}$ which we follow is an integral curve of the vector field $V^{\rm ct}$ we introduced in the last section. Therefore, we take $V^{\rm ct}$ as the direction in the predictor step. Let $\bar{s} \in t\mathcal{D}$, and let

$$\bar{s}_L(t') = \bar{s} + (t' - t)G_s^{-1}(I - \Pi^{\perp})G_s c. \tag{40}$$

As we observed in (37), we have $\bar{s}_L(t') \in t'\mathcal{D}$ for certain interval containing t. We choose step Δt and adopt $\bar{s}_L(t + \Delta t)$ as the result of the predictor step.

[Neighborhood of the central trajectory]

Below we explain the neighborhood of the central trajectory and the corrector step. They are closely related each other since the neighborhood is determined in such a way that the corrector step performs well. In view of this, we proceed as follows. Let $\bar{s} \in t\mathcal{D}$. Recall that $s(t)(= s(\gamma_{\mathcal{P}}(t)))$ is characterized as the optimal solution to the problem (32). The corrector step at s is the Newton step for the point $s(t) \in t\mathcal{D}$ to solve the problem (32).

By using the Newton decrement $\|N_s^{\mathbf{T}^*}(d)\|_s$ for (32), we introduce the neighborhood $\mathcal{N}_t(\beta)$ of the center point $s(t)$ on the slice $t\mathcal{D}$ as follows:

$$\mathcal{N}_t(\beta) \equiv \{s \in t\mathcal{D} \mid \|N_s^{\mathbf{T}^*}(d)\|_s \le \beta.\}$$

The neighborhood $\mathcal{N}(\beta)$ of the central trajectory is determined as

$$\mathcal{N}(\beta) \equiv \cup_{t\in(0,\infty)}\mathcal{N}_t(\beta) = \{s \in \mathrm{Hom}(\mathcal{D}) \mid \|N_s^{\mathbf{T}^*}(d)\|_s \le \beta.\}$$

[Corrector-Step]

After the predictor-step is performed, we have

$$\bar{s}^{\rm P} \equiv \bar{s}_L(t + \Delta t) \in (t + \Delta t)\mathcal{D} \cap \mathcal{N}(\beta).$$

As we discussed above, the corrector-step is the Newton step for the convex optimization problem (32) (with $t := t + \Delta t$) whose optimal solution is $s(t + \Delta t)$. The point $\bar{s}^{\rm P}$ is a feasible solution to this problem, and we apply a single step of the Newton method. This is the corrector step.

Now we are ready to fully describe the implementable algorithm. Below η is a constant to determine accuracy of line search. The numbers 1/4 and 1/2 in Step 1 are picked just for simple presentation and is not essential.

[Implementable algorithm]

1. Let $\beta \le 1/4$, and let $1/2 \le \eta$.

2. Let $\bar{s} \in t\mathcal{D}$ such that $\bar{s} \in \mathcal{N}(\frac{16}{9}\beta^2)$.

3. (Predictor step) Let $\Delta t > 0$ be such that

$$\|N^{\mathbf{T}^*}_{\bar{s}_L(t+\Delta t)}(d)\|_{\bar{s}_L(t+\Delta t)} \in [\beta(1-\eta), \beta],$$

where $\bar{s}_L(\cdot)$ is as defined in (40). Let $\bar{s}^{\mathrm{P}} := \bar{s}_L(t+\Delta t)$.

4. (Corrector step) At $\bar{s}^{\mathrm{P}}$, compute the Newton direction for the corrector-step $N_{\bar{s}_L(t+\Delta t)}(d)$ as above, and let $\bar{s}^+ = \bar{s}^{\mathrm{P}} + N^{\mathbf{T}^*}_{\bar{s}^{\mathrm{P}}}(d)$.

5. $t := t + \Delta t$, $\bar{s} := \bar{s}^+$ and return to step 1.

Finally, we present a theorem on polynomiality of this algorithm.

Theorem 4.3 *The predictor-corrector algorithm above generates the sequence* $\{(t^k, s^k)\}$ *satisfying*

$$s^k \in t^k\mathcal{D} \cap \mathcal{N}(\beta), \quad t^k \geq \left(1 + \frac{\lambda}{2\sqrt{p}}\right)^k.$$

The proof of this algorithm is not shown here but in the appendix. We just outline the idea of the proof. We analyze the sequence $\{s^k/t^k\}$ rather than $\{s^k\}$ itself. $\{s^k/t^k\}$ is a feasible solution to the dual problem (26). We will show that the predictor-corrector algorithm coincides another predictor-corrector algorithm for the dual problem (26) where "the negative centering direction" is utilized as the search direction for the predictor step. Once notifying this fact, the polynomial-time complexity analysis goes through by a rather standard argument established by Nesterov and Nemirovski [42].

In the next sections, we will relate the iteration-complexity of the predictor-corrector algorithm developed in this section to geometrical structure of the feasible regions. Before proceeding, we explain a bit about the main idea. We observed that the central trajectory as the intersection of two submanifolds $\mathcal{P}$ and $\mathrm{Hom}(\mathcal{D})$. In view of s-coordinate, $\mathcal{P}$ is a curved manifold ($\mathrm{Hom}(\mathcal{D})$ is flat), therefore, when we move along the tangent of the trajectory, we will move away from $\mathcal{P}$. We increase the step as long as possible within the range where the corrector step works effectively to bring the iterate closer to $\mathcal{P}$ again (See Fig. 7). This observation leads to the intuition that the steplength would be closely related to how $\mathcal{P}$ is embedded in Ω^* as a curved submanifold. In [45, 46], Ohara analyzed how the embedding curvature is related to performance of a predictor step. In the next section, we develop a concrete analysis of the iteration-complexity of interior-point algorithms based on the embedding curvature.

5 Curvature Integral and Iteration-Complexity of IPM

The goal of this section is to prove a main result of this paper which bridges the iteration-complexity of the predictor-corrector algorithm and the information-geometric structure of the feasible region. The iteration-complexity of the predictor-corrector algorithm is expressed in terms of a curvature integral along the central trajectory $\gamma_{\mathcal{P}}$ involving the directional embedding curvature $H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}(t), \dot\gamma_{\mathcal{P}}(t))$.

In the following, we make the following assumption:

Let $0 \le t_1 \le t_2$ be constants. The step-size $\Delta t(s)$ in [Implementable algorithm] of Section 4.2 taken at $s \in \mathcal{N}(\beta) \cap \{t\mathcal{D} \mid t_1 \le t \le t_2\}$ converges to zero when β goes to zero.

If this assumption is not satisfied, then we conclude that $\gamma_{\mathcal{P}}$ becomes straight (in s-coordinate) at least for a short interval (under the assumption of smoothness). If we assume that the ψ is analytic (so is ψ^*), violation of the assumption implies that $\gamma_{\mathcal{P}}$ is a (half) straight line in s-coordinate leading to the optimal solution directly.

Now we describe the theorem.

Theorem 5.1 *Let $0 < t_1 < t_2$, and let $K(s_1, t_2, \beta)$ be the number of iterations of the predictor-corrector algorithm in Section 4.2 started from a point $s_1 \in \mathcal{N}(\beta) \cap t_1\mathcal{D}$ to find a point in $\mathcal{N}(\beta) \cap t_2\mathcal{D}$. Then we have the following estimate on the number of iterations of the predictor-corrector algorithm.*

1. *If we perform "exact line search" to the boundary of the neighborhood $\mathcal{N}(\beta)$,*

$$K(s_1, t_2, \beta) = \frac{1}{\sqrt{\beta}} \frac{1}{\sqrt{2}} \int_{t_1}^{t_2} \|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^{1/2} dt + \frac{o(1)}{\sqrt{\beta}}.$$

2. *If we perform inexact line search with accuracy η, then,*

$$\begin{aligned} &\frac{1}{\sqrt{\beta}} \frac{1}{\sqrt{2}} \int_{t_1}^{t_2} \|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^{1/2} dt + \frac{o(1)}{\sqrt{\beta}} \\ &\qquad \le K(s_1, t_2, \beta) \le \frac{1}{\sqrt{(1-\eta)\beta}} \frac{1}{\sqrt{2}} \int_{t_1}^{t_2} \|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}, \dot\gamma_{\mathcal{P}})\|^{1/2} dt + \frac{o(1)}{\sqrt{\beta}}. \end{aligned}$$

The theorem is immediately seen from Lemmas 5.3 and 5.4. To prove Lemma 5.3, we need Lemma 5.2.

Before going to prove Lemma 5.2, we will explain the situation with Fig. 8. We denote by $s(t)$ the point on the central trajectory with parameter t written in s-coordinate. Let $\bar s \in \mathcal{N}(c_1) \cap t\mathcal{D}$. This point will be considered as the current iterate in later analysis. The integral curve of $V^{\rm ct}_s$ passing through "the current iterate" $\bar s$ is denoted by $\bar s(\cdot)$. Note that $\bar s(t) = \bar s$. Recall that

$$\bar s_L(t') \equiv \bar s + (t'-t)\dot{\bar s} = \bar s + (t'-t)V^{\rm ct}_{\bar s} \tag{41}$$

is the linear approximation to the integral curve and the predictor-step is taken along this straight line. In order to estimate the step-size Δt in the predictor-step, we need to develop a proper bound for $\|N^{\mathbf{T}^*}_{\bar s_L(t+\Delta t)}(d)\|_{\bar s_L(t+\Delta t)}$.

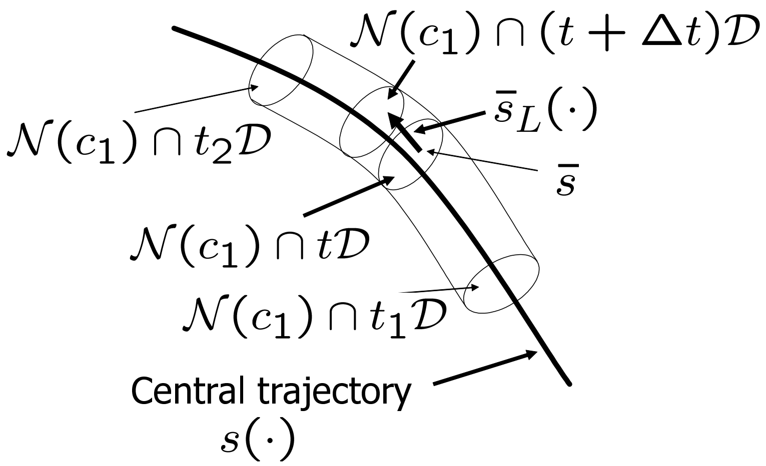

Figure 8: The situation of Lemma 5.2 .

Lemma 5.2 *Let $0 < t_1 < t_2$ and let c_1 be a positive constant, and let*

$$\mathcal{T}(t_1, t_2, c_1) \equiv \cup_{t_1 \le t \le t_2} \mathcal{N}(c_1) \cap t\mathcal{D}.$$

Let

$$R = \max_{s \in \mathcal{T}} \|V_s^{ct}\|_s, \quad M = \max_{s \in \mathcal{T}} \|G_s^{-1/2} \dot{G}_s G_s^{-1/2}\|.$$

($\|\cdot\|$ on the right hand side of M is the usual (conventional) operator norm, and $\dot{G}_s$ is derivative of G_s with respect to $V^{\rm ct}$)

For sufficiently small c_1 such that $c_1 \le 1/100$, there exists M_1, M_2 and M_3, for which the following property holds (see Fig. 8).
"Let $\bar{s} \in \mathcal{T} \cap t\mathcal{D}$, and take the step Δt along $\bar{s}_L(\cdot)$ to obtain the point $\bar{s}_L(t + \Delta t)$. If $\Delta t \le \min(1/(2R), (\log 4)/M)$ and $\bar{s}_L(t + \Delta t) \in \mathcal{T}$, then,

$$\|N^{\mathbf{T}^*}_{\bar{s}_L(t+\Delta t)}(d)\| = \frac{\Delta t^2}{2}\|\ddot{s}(t)\|_{s(t)} + \delta + r_1(\Delta t, s(t), \bar{s}) + r_2(\Delta t, s(t), \bar{s}) + r_3(\Delta t, s(t), \bar{s}),$$

where $\delta \equiv \|N^{\mathbf{T}^*}_{\bar{s}}(d)\|_{\bar{s}}$, $|r_1| \le M_1 \Delta t^3$ and $|r_2| \le M_2 \delta$ and $|r_3| \le M_3(\Delta t^2 + \delta)^2$."

Proof. Our task is to find a proper bound for the Newton decrement $\|N^{\mathbf{T}^*}_{\bar{s}_L(t+\Delta t)}(d)\|$. To this end, we need to analyze "vertical move" and "horizontal move," and need to compare quantities at different points. See Fig. 9. We start with the following observations.

1. Since the metric G is a smooth function and positive definite, and since $\mathcal{T}$ is a compact set, there exists a norm $\|\cdot\|_-$ and $\|\cdot\|_+$ with which the local norm $\|\cdot\|_s$ at each point in $\mathcal{T}$ is bounded from below and above, respectively, i.e., for any $V \in \mathbf{E}^*$ and $s \in \mathcal{T}$, we have $\|V\|_- \le \|V\|_s \le \|V\|_+$.

2. At any $\bar{s} \in \mathcal{T} \cap t\mathcal{D}$ and $V \in \mathbf{E}^*$,

$$(1 - R\Delta t)\|V\|_{\bar{s}} \leq \|V\|_{\bar{s}_L(t+\Delta t)} \leq \frac{1}{1 - R\Delta t}\|V\|_{\bar{s}} \quad (\bar{s}_L(t) = \bar{s}(t) = \bar{s}).$$

(Proof) To see this, we observe that $\bar{s}_L(t + \Delta t)$ is contained in the Dikin ellipsoid of radius $R\Delta t$ at s. Indeed, this is the case, since

$$\bar{s}_L(t + \Delta t) = \bar{s} - (\Delta t\|V_{\bar{s}}^{\rm ct}\|_{\bar{s}})\frac{V_{\bar{s}}^{\rm ct}}{\|V_{\bar{s}}^{\rm ct}\|_{\bar{s}}}$$

Since $-\frac{V_{\bar{s}}^{\rm ct}}{\|V_{\bar{s}}^{\rm ct}\|_{\bar{s}}}$ is the displacement vector to the surface of the Dikin ellipsoid with radius 1 (centered at $\bar{s}$), we see that taking the step Δt means that moving to the surface of the Dikin ellipsoid with radius $\Delta t\|V_{\bar{s}}^{\rm ct}\|_{\bar{s}}$ $(\leq R\Delta t)$.

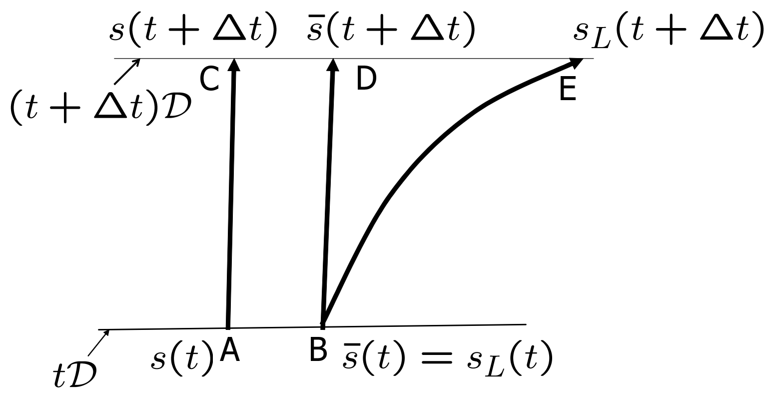

Figure 9: The predictor-corrector algorithm .

3. Let $\bar{s} \in \mathcal{T} \cap t\mathcal{D}$. Recall $\|N_{\bar{s}}^{\mathbf{T}^*}(d)\|_{\bar{s}}$ is the Newton decrement at $\bar{s}$ for the point $s(t)$ in $t\mathcal{D}$ of the central trajectory. We have the following result:

$$\|N_{\bar{s}}^{\mathbf{T}^*}(d)\|_{\bar{s}}(1 - 8\|N_{\bar{s}}^{\mathbf{T}^*}(d)\|_{\bar{s}}) \leq \|s(t) - \bar{s}\|_{\bar{s}} \leq \|N_{\bar{s}}^{\mathbf{T}^*}(d)\|_{\bar{s}}(1 + 8\|N_{\bar{s}}^{\mathbf{T}^*}(d)\|_{\bar{s}})$$

and

$$\|s(t) - \bar{s}\|_{\bar{s}}(1 - 22\|s(t) - \bar{s}\|_{\bar{s}}) \leq \|N_{\bar{s}}^{\mathbf{T}^*}(d)\|_{\bar{s}} \leq \|s(t) - \bar{s}\|_{\bar{s}}(1 + 22\|s(t) - \bar{s}\|_{\bar{s}}).$$

(Proof) This readily follows from Proposition 3.7 and that $c_1 \leq 1/100$.

4. There exists a positive constant M' such that for any $\bar{s} \in t\mathcal{D} \cap \mathcal{T}$,

$$\|\ddot{s}(t) - \ddot{\bar{s}}(t)\|_{\bar{s}(t)} \leq M'\|N_{\bar{s}(t)}^{\mathbf{T}^*}(d)\|_{\bar{s}(t)}.$$

(Proof) Recall that $\dot{\bar{s}} = V_{\bar{s}}^{\rm ct}$. Since

$$\ddot{\bar{s}}(t) = \ddot{s}(\bar{s}(t)) = \nabla^*_{V_{\bar{s}(t)}^{\rm ct}} V_{\bar{s}(t)}^{\rm ct},$$

we see that $\ddot{s}(\cdot)$ is a smooth function of $\bar{s}$ (see (39)). Therefore, there is a positive constant M'' with which the following bound for any $s_1, s_2 \in \mathcal{T}$,

$$\|\ddot{s}(s_1) - \ddot{s}(s_2)\|_+ \leq M''\|s_1 - s_2\|_-.$$

Then it readily follows from the claim 1 that

$$\|\ddot{s}(s_1) - \ddot{s}(s_2)\|_{s_2} \leq M''\|s_1 - s_2\|_{s_2}.$$

We let $s_1 = s(t)$ and $s_2 = \bar{s}(t)$ in the above relation and use the claim 3. Then the statement readily follows.

5. We have

$$\|N^{\mathbf{T}^*}_{\bar{s}(t+\Delta t)}(d)\|_{\bar{s}(t+\Delta t)} \leq \exp(M\Delta t)\|N^{\mathbf{T}^*}_{\bar{s}(t)}(d)\|_{\bar{s}(t)} \leq 2\|N^{\mathbf{T}^*}_{\bar{s}(t)}(d)\|_{\bar{s}(t)}.$$

(Proof) Let $t' \in \mathbf{R}$. The Newton decrement is written as (see (24)):

$$\|N^{\mathbf{T}^*}_{\bar{s}(t')}(d)\|_{\bar{s}(t')} = \sqrt{(b - Ax(\bar{s}(t'))(AG_{\bar{s}(t')}A^\top)^{-1}(b - Ax(\bar{s}(t'))}.$$

We focus on

$$(b - Ax(\bar{s}(t')))^\top(AG_{\bar{s}(t')}A^\top)^{-1}(b - Ax(\bar{s}(t'))),$$

and integrate it along the trajectory. The curve $\bar{s}(\cdot)$ is the integral curve of V^{ct} which is the central trajectory for the problem (38) with $d' = x(\bar{s})$. Therefore, for any t' on the integral curve, we have $x(\bar{s}(t')) - x(\bar{s}) \in \mathbf{T}$ and hence $A(x(\bar{s}(t')) - x(\bar{s})) = 0$. This means that $Ax(\bar{s}(t'))$ is the constant vector $Ax(\bar{s})$ independent of t'. Therefore,

$$\|N^{\mathbf{T}^*}_{\bar{s}(t')}(d)\|^2_{\bar{s}(t)} = (b-Ax(\bar{s}))^\top(AG_{\bar{s}(t')}A^\top)^{-1}(b-Ax(\bar{s})) = (b-Ax(\bar{s}))^\top(AG_{\bar{s}(t')}A^\top)^{-1}(b-Ax(\bar{s})).$$

Since we have $\|G_s^{-1/2}\dot{G}_s G_s^{-1/2}\| \leq M$,

$$\begin{aligned}
&\frac{d}{dt}\|N_{\bar{s}_L(t')}(d)^{\mathbf{T}^*}\|^2_{\bar{s}_L(t+\Delta t)} \\
&= \left|\frac{d}{dt}(b - Ax(\bar{s}))^\top(AG_{\bar{s}(t')}A^\top)^{-1}(b - Ax(\bar{s}))\right| \\
&= |(b - Ax(\bar{s}))^\top(AG_{\bar{s}(t')}A^\top)^{-1}G_s^{-1}\dot{G}_sG_s^{-1}(AG_{\bar{s}(t')}A^\top)^{-1}(b - Ax(\bar{s}))| \\
&\leq |(b - Ax(\bar{s}))^\top(AG_{\bar{s}(t')}A^\top)^{-1}AG^{1/2}_{\bar{s}(t')}G_s^{-1/2}\dot{G}_sG_s^{-1/2}G^{1/2}_{\bar{s}(t')}A^T(AG_{\bar{s}(t')}A^\top)^{-1}(b - Ax(\bar{s}))| \\
&\leq |(b - Ax(\bar{s}))^\top(AG_{\bar{s}(t')}A^\top)^{-1}AG^{1/2}_{\bar{s}(t')}G_s^{-1/2}\dot{G}_sG_s^{-1/2}G^{1/2}_{\bar{s}(t')}A^T(AG_{\bar{s}(t')}A^\top)^{-1}(b - Ax(\bar{s}))| \\
&\leq \|G_s^{-1/2}\dot{G}_sG_s^{-1/2}\|\|G^{1/2}_{\bar{s}(t')}A^T(AG_{\bar{s}(t')}A^\top)^{-1}(b - Ax(\bar{s}))\|^2 \\
&\leq M(b - Ax(\bar{s}))^\top(AG_{\bar{s}(t')}A^\top)^{-1}(b - Ax(\bar{s})) \\
&= M\|N^{\mathbf{T}^*}_{\bar{s}_L(t')}(d)\|^2_{\bar{s}_L(t')}
\end{aligned}$$

Therefore, we have

$$\|N^{\mathbf{T}^*}_{\bar{s}_L(t+\Delta t)}(d)\|^2_{\bar{s}_L(t)} \leq \exp(M\Delta t)\|N^{\mathbf{T}^*}_{\bar{s}_L(t)}(d)\|^2_{\bar{s}_L(t)}$$

and hence if $\Delta t \leq (\log 4)/M$, then,

$$\|N_{\bar{s}(t+\Delta t)}^{\mathbf{T}^*}(d)\|_{\bar{s}(t+\Delta t)} \leq 2\|N_{\bar{s}(t)}^{\mathbf{T}^*}(d)\|_{\bar{s}(t)}.$$

6. Let w, z be vectors of the same dimension. If $\|z\| \leq r$, then,

$$|\|w+z\| - \|w\|| \leq r.$$

(Proof) Easily follows from the triangular inequality.

Now we are ready to prove Lemma 5.2. We have

$$\bar{s}(t+\Delta t) - \bar{s}_L(t+\Delta t) = \frac{(\Delta t)^2}{2}\ddot{s}(\bar{s}(t)) + r_1'(\Delta t, \bar{s}),$$

where $\|r_1'\|_+ \leq M_1'\Delta t^3$ (Fig. 9). On the other hand, due to the claims 3 and 5, we have

$$s(t+\Delta t) - \bar{s}(t+\Delta t) = r_2'(\Delta t, s, \bar{s}),$$

where $\|r_2'\|_+ \leq M_2'\delta$. Due to the claim 4, we have

$$\frac{(\Delta t)^2}{2}\ddot{s}(\bar{s}(t)) = \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) + r_3'(s, \bar{s}),$$

where $\|r_3'\|_+ \leq M_3'\Delta t^2\delta$. Putting these estimates altogether, we have,

$$\begin{aligned} s(t+\Delta t) - \bar{s}_L(t+\Delta t) &= (\bar{s}(t+\Delta t) - \bar{s}_L(t+\Delta t)) + (s(t+\Delta t) - \bar{s}(t+\Delta t)) \\ &= \left(\frac{(\Delta t)^2}{2}\ddot{s}(\bar{s}(t)) + r_1'(\Delta t, \bar{s})\right) + r_2'(\Delta t, s, \bar{s}) \\ &= \left(\frac{(\Delta t)^2}{2}\ddot{s}(s(t)) + r_3'(s, \bar{s})\right) + r_1'(\Delta t, \bar{s}) + r_2'(\Delta t, s, \bar{s}). \end{aligned} \tag{42}$$

Therefore, for some constant $M_4' > 0$, we have

$$\|s(t+\Delta t) - \bar{s}_L(t+\Delta t)\|_+ \leq M_4'(\Delta t^2 + \delta).$$

Now it follows from the claim 3 that

$$\|N_{\bar{s}_L(t+\Delta t)^{\mathbf{T}^*}}(d)\|_{\bar{s}_L(t+\Delta t)} = \|s(t+\Delta t) - \bar{s}_L(t+\Delta t)\|_{\bar{s}_L(t+\Delta t)} + r_4',$$

where

$$|r_4'| \leq 44\|s(t+\Delta t) - \bar{s}_L(t+\Delta t)\|_+^2 \leq 44M_4'^2(\Delta t^2 + \delta)^2.$$

Applying the claim 6 to (42), we have

$$\|s(t+\Delta t) - \bar{s}_L(t+\Delta t)\|_{\bar{s}_L(t+\Delta t)} = \left\|\frac{(\Delta t)^2}{2}\ddot{s}(s(t))\right\|_{\bar{s}_L(t+\Delta t)} + \hat{r}_3(\bar{s}(t), s(t)) + \hat{r}_1(\Delta t, \bar{s}) + \hat{r}_2(\Delta t, s, \bar{s}),$$

where

$$\begin{aligned}
\hat{r}_3 &= \left\| \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) + r'_1(\Delta t,\bar{s}) + r'_2(\Delta t, s,\bar{s}) + r'_3(\bar{s}(t), s(t)) \right\|_{\bar{s}_L(t+\Delta t)} \\
&\quad - \left\| \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) + r'_1(\Delta t,\bar{s}) + r'_2(\Delta t, s,\bar{s}) \right\|_{\bar{s}_L(t+\Delta t)} \\
\hat{r}_2 &= \left\| \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) + r'_1(\Delta t,\bar{s}) + r'_2(\Delta t, s,\bar{s}) \right\|_{\bar{s}_L(t+\Delta t)} - \left\| \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) + r'_1(\Delta t,\bar{s}) \right\|_{\bar{s}_L(t+\Delta t)} \\
\hat{r}_1 &= \left\| \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) + r'_1(\Delta t,\bar{s}) \right\|_{\bar{s}_L(t+\Delta t)} - \left\| \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) \right\|_{\bar{s}_L(t+\Delta t)}
\end{aligned}$$

and $\hat{r}_1 \le M'_1 \Delta t^3$, $\hat{r}_2 = M'_2\delta$, $\hat{r}_3 \le M'_3 \Delta t^2 \delta$, and hence,

$$\begin{aligned}
\|N_{\bar{s}_L(t+\Delta t)}\|_{\bar{s}_L(t+\Delta t)} &= \|s(t+\Delta t) - \bar{s}_L(t+\Delta t)\|_{\bar{s}_L(t+\Delta t)} + r'_4 \\
&= \left\| \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) \right\|_{\bar{s}_L(t+\Delta t)} + \hat{r}_3(\bar{s}(t), s(t)) + \hat{r}_1(\Delta t,\bar{s}) + \hat{r}_2(\Delta t, s, \bar{s}) + r'_4.
\end{aligned}$$

Finally, due to Theorem 3.1 and the claims 2 and 3, we have

$$\begin{aligned}
\frac{1}{2}(1-2\delta)\|\ddot{s}(s(t))\|_s &\le (1 - R\Delta t)(1-2\delta)\|\ddot{s}(s(t))\|_s \\
&\le \|\ddot{s}(s(t))\|_{\bar{s}_L(t+\Delta t)} \\
&\le \frac{1}{1-R\Delta t}\frac{1}{1-2\delta}\|\ddot{s}(s(t))\|_s \le \frac{2}{1-2\delta}\|\ddot{s}(s(t))\|_s.
\end{aligned}$$

This implies that

$$\|N_{\bar{s}_L(t+\Delta t)}\|_{\bar{s}_L(t+\Delta t)} = \left\| \frac{(\Delta t)^2}{2}\ddot{s}(s(t)) \right\|_{\bar{s}_L(t+\Delta t)} + \hat{r}_3(\bar{s}(t), s(t)) + \hat{r}_1(\Delta t,\bar{s}) + \hat{r}_2(\Delta t, s, \bar{s}) + r'_4 + r'_5,$$

where $|r'_5| \le M_5 \Delta t^2(\Delta t + \delta)$. This completes the proof. ∎

Lemma 5.3 *Let $K(s_1, t_2, \beta)$ be the number of iterations of the predictor-corrector algorithm in Section 4.2 from a point $s_1 \in \mathcal{N}(\beta)$ such that $s \in t_1\mathcal{D}$ and $\|N^{\mathbf{T}^*}_{s_1}(d)\|_{s_1} \le \frac{16\beta^2}{9}$. Then we have the following estimate on the number of iterations of the predictor-corrector algorithm:*

$$\frac{1}{\sqrt{\beta}}\frac{1}{\sqrt{2}}\int_{t_1}^{t_2} \|\ddot{s}(t)\|^{1/2}_{s(t)}dt + \frac{o(1)}{\sqrt{\beta}} \le K(x_1, t_2, \beta) \le \frac{1}{\sqrt{(1-\eta)\beta}}\frac{1}{\sqrt{2}}\int_{t_1}^{t_2} \|\ddot{s}(t)\|^{1/2}_{s(t)}dt + \frac{o(1)}{\sqrt{\beta}}.$$

Proof. Let $\{s^k\}$ be the generated sequence, and assume that $s^k \in t^k\mathcal{D}$. Then, due to Lemma 5.2, we have

$$\begin{aligned}
\beta^{k+1} &= \|N^{\mathbf{T}^*}_{s^k_L(t^k+\Delta t^{k+1})}(d)\|_{s^k_L(t^k+\Delta t^{k+1})} \\
&= \frac{\Delta(t^k)^2}{2}\|\ddot{s}(t^k)\|_{s(t^k)} + \|N^{\mathbf{T}^*}_{s^k}(d)\|_{s^k} + r_1(\Delta t^k, \|N^{\mathbf{T}^*}_{s^k}(d)\|_{s^k}) \\
&\quad + r_2(\Delta t^k, \|N^{\mathbf{T}^*}_{s^k}(d)\|_{s^k}) + r_3(\Delta t^k, \|N^{\mathbf{T}^*}_{s^k}(d)\|_{s^k}),
\end{aligned}$$

where $|r_1^k| \le M_1(\Delta t^k)^3$ and $|r_2^k| \le M_2\delta^k$ and $|r_3| \le M_3((\Delta t^k)^2 + \delta^k)^2$ and M_1, M_2, M_3 are the constants appear in Lemma 5.2 and do not depend on k. We use abbreviation that $r_1^k = r_1(\Delta t^k, \|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k})$ and $r_2^k = r_2(\Delta t^k, \|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k})$, etc. Let

$$w^k = \beta^{k+1} - r_2^k - \|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k}.$$

Observe that w^k is strictly positive if β is sufficiently small, since $\beta^{k+1} \ge (1-\eta)\beta$ while r_2^k and $\|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k}$ is $O(\beta^2)$. Then, we have, when β is sufficiently small, for each $k = 1, \dots, K(s_1, t_2, \beta)$,

$$\sqrt{\frac{\|\ddot{s}(t^k)\|_{s(t^k)}}{2}(\Delta t^k)^2 + r_1^k + r_3^k} = \sqrt{w^k}.$$

This implies that

$$\frac{\Delta t}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2} - \sqrt{|r_1^k|} - \sqrt{|r_3^k|} \le \sqrt{w^k} \le \frac{\Delta t}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2} + \sqrt{|r_1^k|} + \sqrt{|r_3^k|}.$$

Since $|r_3^k| \le M_3((\Delta t^k)^2 + \|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k})^2$, we have

$$\frac{\Delta t}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2} - \sqrt{|r_1^k|} - \sqrt{M_3}(\Delta t^k)^2 \le \sqrt{w^k} + \sqrt{M_3}\|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k}$$

and

$$\sqrt{w^k} - \sqrt{M_3}\|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k} \le \frac{\Delta t^k}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2} + \sqrt{|r_1^k|} + \sqrt{M_3}(\Delta t^k)^2.$$

Now, we take the summation for all $k = 1, \dots, K(s_1, t_2, \beta)$. Let $\Delta t_{\max}(\beta)$ be $\max_k \Delta t^k$.

Since $r_1^k = O((\Delta t^k)^3)$, we have

$$\int_{t_1}^{t_2} \frac{1}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2} dt - M'\sqrt{\Delta t_{\max}(\beta)} \le \sum_k \left(\frac{\Delta t^k}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2} - \sqrt{|r_1^k|} - \sqrt{M_3}(\Delta t^k)^2\right)$$

and

$$\sum_k \left(\frac{\Delta t}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2} + \sqrt{|r_1^k|} + \sqrt{M_3}(\Delta t^k)^2\right) \le \int_{t_1}^{t_2} \frac{1}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2} dt + M'\sqrt{\Delta t_{\max}(\beta)}$$

where M' is a positive constant. Therefore, we have

$$\int_{t_1}^{t_2} \frac{1}{\sqrt{2}}\|\ddot{s}(t)\|_{s(t^k)}^{1/2} dt - M'\sqrt{\Delta t_{\max}(\beta)} \le \sum_{i=1}^{K(s_1,t_2,\beta)} \left(\sqrt{w^k} + \sqrt{M_3}\|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k}\right) \tag{43}$$

and

$$\sum_{i=1}^{K(s_1,t_2,\beta)} \left(\sqrt{w^k} - \sqrt{M_3}\|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k}\right) \le \int_{t_1}^{t_2} \frac{1}{\sqrt{2}}\|\ddot{s}(t)\|_{s(t)}^{1/2} dt + M'\sqrt{\Delta t_{\max}(\beta)} \tag{44}$$

for sufficiently small β (and any s_1), where M is a positive constant.

Now, we further analyze $\sum_{i=1}^{K(s_1,t_2,\beta)}(\sqrt{w^k} \pm \sqrt{M_3}\|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k})$. Recall that $\beta^k \in [(1-\eta)\beta, \beta]$, $r_2^k = O(\delta^k)$ and $\delta^k = \|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k} = O(\beta^2)$. Then,

$$\begin{aligned}\sqrt{(1-\eta)\beta}(1-O(\sqrt{\beta})) &\leq \sqrt{w^k} \pm \sqrt{M_3}\|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k} = \sqrt{\beta^k - r_2^k - \|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k}} \pm \sqrt{M_3}\|N_{s^k}^{\mathbf{T}^*}(d)\|_{s^k} \\ &\leq \sqrt{\beta}(1+O(\sqrt{\beta})). \qquad (45)\end{aligned}$$

Taking summation from $k = 1, \ldots, K(s_1, t_2, \beta)$, we have

$$\sqrt{\beta(1-\eta)}K(s_1,t_2,\beta)[1-O(\sqrt{\beta})] \leq \sum_{i=1}^{K(s_1,t_2,\beta)} \sqrt{w^k} \leq \sqrt{\beta}K(s_1,t_2,\beta)[1+O(\sqrt{\beta})]. \qquad (46)$$

Combining (43), (44), (45) and (46), we obtain, finally,

$$\begin{aligned}\sqrt{\beta(1-\eta)}\sum_{k=1}^{K(s_1,t_2,\beta)}\left[1-O(\sqrt{\beta})\right] - M'\sqrt{\Delta t_{\max}(\beta)} &\leq \int_{t_1}^{t_2}\frac{1}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2}dt \\ &\leq \sqrt{\beta}\sum_{k=1}^{K(s_1,t_2,\beta)}\left[1+O(\sqrt{\beta})\right] + M'\sqrt{\Delta t_{\max}(\beta)}.\end{aligned}$$

Since $\Delta t_{\max}(\beta) \to 0$ when $\beta \to 0$, we have

$$\frac{1}{\sqrt{\beta}}\int_{t_1}^{t_2}\frac{1}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t)}^{1/2}dt + \frac{o(1)}{\sqrt{\beta}} \leq K(s_1,t_2,\beta)$$

and

$$K(s_1,t_2,\beta) \leq \frac{1}{\sqrt{\beta(1-\eta)}}\int_{t_1}^{t_2}\frac{1}{\sqrt{2}}\|\ddot{s}(t^k)\|_{s(t^k)}^{1/2}dt + \frac{o(1)}{\sqrt{\beta}}.$$

■

Lemma 5.4 *We have*

$$\ddot{s} = \nabla^*_{\dot{s}}\dot{s} = H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}}).$$

Proof. For simplicity, we let $G := G_x$. Then,

$$\ddot{s} = -\frac{d}{dt}(G\dot{x}) = -\dot{G}\dot{x} - G\ddot{x}.$$

It follows from the definition (36) of $\dot{x}$ that

$$\begin{aligned}G\ddot{x} &= -G\frac{d}{dt}(I-\Pi^\perp)G^{-1}c \\ &= G\dot{\Pi}^\perp G^{-1}c + G(I-\Pi^\perp)G^{-1}\dot{G}G^{-1}c.\end{aligned}$$

We derive expression for $\dot{\Pi}^\perp$ below. Since

$$G^{-1}A^\top\left(\frac{d}{dt}(AG^{-1}A^\top)^{-1}\right)A = -G^{-1}A^\top(AG^{-1}A^\top)^{-1}\left(\frac{d}{dt}(AG^{-1}A^\top)\right)(AG^{-1}A^\top)^{-1}A = \Pi^\perp G^{-1}\dot{G}\Pi^\perp,$$

we obtain

$$\dot{\Pi}^{\perp} = \frac{d}{dt}\left(G^{-1}A^{\mathsf{T}}(AG^{-1}A^{\mathsf{T}})^{-1}A\right) = \Pi^{\perp}G^{-1}\dot{G}\Pi^{\perp} - G^{-1}\dot{G}\Pi^{\perp} = (\Pi^{\perp} - I)G^{-1}\dot{G}\Pi^{\perp}.$$

Now we have

$$G\ddot{x} = G(I-\Pi^{\perp})G^{-1}\dot{G}(I-\Pi^{\perp})G^{-1}c \quad \text{and} \quad \dot{G}\dot{x} = -\dot{G}(I-\Pi^{\perp})G^{-1}c.$$

Then it immediately follows that

$$\ddot{s} \;=\; -G\ddot{x} - \dot{G}\dot{x} = G\Pi^{\perp}G^{-1}\dot{G}(I-\Pi^{\perp})G^{-1}c.$$

Now,

$$\dot{G} = \frac{\partial^3\psi}{\partial x^3}\dot{x}.$$

Let $K = G^{-1} = \partial^2\psi^*/\partial s^2$. Then, we have $\dot{G} = -G\dot{K}G$. Taking account that $G\Pi^{\perp}G^{-1}$ is the orthogonal projection onto $T_x(\mathcal{P})^{\perp}$ in s-coordinate and that $\dot{s} = G(I-\Pi^{\perp})G^{-1}c$ (see (38)), we obtain

$$\begin{aligned}\nabla^*_{\dot{s}}\dot{s} &= \ddot{s} = -G\Pi^{\perp}\dot{K}G(I-\Pi^{\perp})G^{-1}c = -G\Pi^{\perp}\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s}) \\ &= -(G\Pi^{\perp}G^{-1})G\Pi^{\perp}\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s}) = (G\Pi^{\perp}G^{-1})\nabla^*_{\dot{s}}\dot{s} = H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}}). \qquad (47)\end{aligned}$$

This completes the proof. ■

In the end of this section, we prove the following result that the $\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}})\|^{1/2}$ is bounded by $\sqrt{p}/t$. The result implies the bound

$$\int_{t_1}^{t_2}\frac{1}{\sqrt{2}}\|H_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}})\|^{1/2}dt \le \sqrt{p}\log\frac{t_2}{t_1},$$

which is naturally expected from the standard complexity analysis of interior-point algorithms.

Proposition 5.5 *We have*

$$\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}})\| \le \frac{2p}{t^2}.$$

Proof. Since $G(V,\Pi^{\perp}V) = G(\Pi^{\perp}V,\Pi^{\perp}V)$ holds for $V\in T_x\Omega$, we have

$$\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}})\|_s^2 = \|G_x\Pi^{\perp}\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s})\|_s^2 = \langle G_x\Pi^{\perp}\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s}),\Pi^{\perp}\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s})\rangle = \langle G_x\Pi^{\perp}\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s}),\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s})\rangle.$$

By using the relation (9) (regarding ψ^* as the self-concordant function), we have

$$\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}})\|^2 = \|G_x\Pi^{\perp}\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s})\|_s^2 \le 2\|G_x\Pi^{\perp}\frac{\partial^3\psi^*}{\partial s^3}(\dot{s},\dot{s})\|_s\|\dot{s}\|_s\|\dot{s}\|_s.$$

Therefore,

$$\|H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}},\dot{\gamma}_{\mathcal{P}})\| \le 2\|\dot{s}\|_s^2.$$

Now, since $s = tc - A^{\mathsf{T}}y$ for some y and $G_s^{-1}(I - \Pi^{\perp})G_s A^{\mathsf{T}}y = 0$, we have

$$t\dot{s} = G_s^{-1}(I - \Pi^{\perp})G_s s = -G_x(I - \Pi^{\perp})\frac{\partial \psi^*}{\partial s}.$$

The norm of this vector is bounded as follows:

$$\left\| G_x(I - \Pi^{\perp})\frac{\partial \psi^*}{\partial s} \right\|_s \le \left\| \frac{\partial \psi^*}{\partial s} \right\|_s = \sqrt{p},$$

where the last equation follows by letting $S := s$ in (15) and by using (14). This completes the proof. ■

6 Linear Programming

In this section, we focus on classical linear programming [12]. Let us consider the dual pair of linear programs:

$$\min \ c^{\mathsf{T}}x \text{ s.t. } Ax = b, \ x \ge 0,$$

and

$$\max \ b^{\mathsf{T}}y \text{ s.t. } c - A^{\mathsf{T}}y = s, \ s \ge 0,$$

where $A \in R^{m\times n}$, $c \in \mathbf{R}^n$, $b \in \mathbf{R}^m$. We assume that the rows of A is linearly independent. It is easy to see that the problem fits into the general setting in (1) and (2), if we take $\Omega = \Omega^* = \mathbf{R}^n_+$ and take d satisfying $Ad = -b$. We will consider the situation where we choose $\psi(x) = -\sum_{i=1}^n \log x_i$ which is an n-normal barrier.

Let

$$\bar{\chi}_A = \max_B \|A_B^{-1}A\|,$$

where B is the set of indices such that A_B is nonsingular. Furthermore, let

$$\bar{\chi}_A^* = \inf_D \ \bar{\chi}_{AD},$$

where D is the positive definite diagonal matrix. The quantity $\bar{\chi}_A$ is the condition number of the coefficient matrix A studied in, for example, [16, 24, 54, 62, 63]. This quantity plays an important role in the layered-step interior-point algorithm by Vavasis and Ye [71], and the subsequent analysis by Monteiro and Tsuchiya [39]. The quantity $\bar{\chi}_A^*$ is a scaling-invariant version of $\bar{\chi}_A$ introduced in [39]. If A is integral, then, $\bar{\chi}_A$ is bounded by $2^{O(L_A)}$, where L_A is the input size of A.

The main goal of this section is to establish a bound on "the total curvature of the central trajectory." Before going to prove the results, we introduce a few notations here. Given two vectors u and v, we denote the elementwise product as $u \circ v$. The unit element of this product is the vector of all ones and denoted by e. The inverse and square root of u is denoted by u^{-1} and $u^{1/2}$, respectively. u^{-1} is the vector whose elements are reciprocal of the elements of u. This elementwise product is the

Euclidean Jordan algebra associated with the cone $\mathbf{R}^n_+$. As to the order of operations, we promise that the product $\circ$ is weaker than the ordinary product of matrix and vectors, i.e., $Ax \circ y$, say, is interpreted as $(Ax) \circ y$ and not as $A(x \circ y)$. Let $x, s \in \mathbf{R}^n_{++}$. We have $x(s) = x^{-1}$, $s(x) = s^{-1}$, $G_x = (\mathrm{diag}(x))^{-2}$, $G_s = (\mathrm{diag}(s)^{-2})$. We define the projection matrix Q as follows:

$$Q(s) = G_s^{1/2} A^\top (A G_s A^\top)^{-1} A G_s^{1/2}.$$

We also use the notation $\|\cdot\|_2$ for the ordinary Euclidean norm defined by $\|u\|_2 = \sqrt{(u|u)} = \sqrt{\sum_i u_i^2}$ for a vector u, say. $\|\cdot\|$ means the norm in terms of the Riemannian metric.

In the analysis of the total curvature integral, we invoke the following theorem obtained by Monteiro and Tsuchiya [39].

Theorem 6.1 *Let $x(\nu), s(\nu), y(\nu)$ be the point of the central trajectory with parameter ν which is defined as the unique solution to the following system of equations.*

$$\begin{aligned} x \circ s &= \nu e, \\ Ax &= b, \\ c - A^\top y &= s, \\ x &\geq 0,\ s \geq 0. \end{aligned}$$

Then we have

$$\int_0^\infty \frac{\sqrt{\nu}\|\dot{x} \circ \dot{s}\|_2^{1/2}}{\nu} d\nu \leq O(n^{3.5} \log(\bar{\chi}_A^* + n)), \tag{48}$$

where differentiation is with respect to ν.

Proposition 6.2

$$\int_0^\infty \frac{\|(I - Q)e \circ Qe\|_2^{1/2}}{t} dt = O(n^{3.5} \log(\bar{\chi}_A^* + n))$$

Proof. We observe that

$$\nu x^{-1} \circ \dot{x} = (I - Q)e, \quad \nu s^{-1} \circ \dot{s} = Qe.$$

where differentiation is with respect to ν in (48). This implies that

$$\nu \dot{x} \circ \dot{s} = (I - Q)e \circ Qe$$

and hence

$$\int_0^\infty \frac{\sqrt{\nu}\|\dot{x} \circ \dot{s}\|_2^{1/2}}{\nu} d\nu = \int_0^\infty \frac{\|(I - Q)e \circ Qe\|_2^{1/2}}{\nu} d\nu.$$

We make change of variables $t = \nu^{-1}$ in the integral. Then the lemma immediately follows. ■

In the following, we let

$$h_{PD}(t) = \frac{\|(I - Q(t))e \circ Q(t)e\|_2}{t^2},$$

where $Q(t)$ is Q defined with $Q(\gamma_{\mathcal{P}}(t)) = Q(\gamma_{\mathcal{D}}(t))$. Now we are ready to prove the main results in this section.

Theorem 6.3 (Total curvature of the central trajectory in the case of "classical linear programming")

If $\Omega = \mathbf{R}^n_+$ and $\psi = \sum_{i=1}^n \log x_i$, then the total curvature of the central trajectory is finite (exists in the improper sense) and is bounded as follows:

$$\int_0^\infty \frac{1}{\sqrt{2}} \| H^*_{\mathcal{P}}(\dot{\gamma}_{\gamma P}, \dot{\gamma}_{\mathcal{P}})\|^{1/2} dt \le \int_0^\infty h_{PD}(t)^{1/2} dt = O(n^{3.5} \log(\bar{\chi}^*_A + n)).$$

Specifically, if A is integral, then

$$\int_0^\infty \frac{1}{\sqrt{2}} \| H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\|^{1/2} dt = O(n^{3.5} L_A),$$

where L_A is the input bit size of A, and in particular, if A is a 0-1 matrix, then

$$\int_0^\infty \frac{1}{\sqrt{2}} \| H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\|^{1/2} dt = O(n^{4.5} m).$$

Proof. We show that

$$\| H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\| \le \|(I-Q)e \circ Qe\|_2 t^2.$$

To this end, we use (47). First observe that

$$\frac{\partial^3 \psi^*}{\partial s_i \partial s_j \partial s_k} = -2 s_i^{-1} s_j^{-1} s_k^{-1} \delta_{ijk}, \tag{49}$$

where $\delta_{ijk} = 1$ iff $i = j = k$ and otherwise zero, and

$$\Pi^\perp = G_s^{1/2} Q G_s^{-1/2}. \tag{50}$$

On the other hand, for any $u \in \mathbf{R}^m$, we have

$$(I-Q) G_s^{1/2} A^\mathsf{T} u = (I-Q)\mathrm{diag}(x(t)) A^\mathsf{T} u = (I-Q)(x(t) \circ (A^\mathsf{T} u)) = (I-Q)(x(t) \circ (A^\mathsf{T} u)) = 0. \tag{51}$$

Therefore,

$$(s^{-1}) \circ \dot{s} = (I-Q)(x \circ c) = (I-Q)(x(t) \circ (t^{-1}(tc - A^\mathsf{T} y(t)))) = t^{-1}(I-Q)e. \tag{52}$$

Taking account of (49), (50), (51) and (52) and (47), we obtain, in s-coordinate,

$$H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}}) = -2t^{-2} \Pi^\perp G_s^{1/2} [(I-Q)e \circ (I-Q)e]. \tag{53}$$

Now we show that

$$\| H^*_{\mathcal{P}}(\dot{\gamma}_{\mathcal{P}}, \dot{\gamma}_{\mathcal{P}})\| = 2t^{-2} \|Q((I-Q)e \circ (I-Q)e)\|_2 \le 2t^{-2} \|Qe \circ (I-Q)e\|_2 = 2h_{PD}. \tag{54}$$

The theorem readily follows from Theorem 6.2 and (54). The first equality relation in (54) is just by definition. We prove the second inequality. Since $Q(I-Q)e = (I-Q)Qe = 0$, we have

$$Q((I-Q)e \circ (I-Q)e) = -Q((I-Q)e \circ Qe) \text{ and } (I-Q)(Qe \circ Qe) = -(I-Q)((I-Q)e \circ Qe).$$

This implies that

$$Q((I-Q)e\circ(I-Q)e)+(I-Q)(Qe\circ Qe) = -Q((I-Q)e\circ Qe)-(I-Q)((I-Q)e\circ Qe) = -Qe\circ(I-Q)e.$$

We take $\|\cdot\|_2$ norm of the both relations and by using $Q(I-Q)=0$, we obtain (54) as we desire. ■

From the proof of Theorem 6.2, we have the following theorem which completely characterizes the primal-dual curvature $h_{PD}(t)$ [39] in terms of information geometry.

Theorem 6.4 *Let $\gamma_{\mathcal{P}}(t)$ and $\gamma_{\mathcal{D}}(t)$ be the points on the primal and the dual central trajectory with parameter t. We have the following Pythagoras relation among the primal embedding curvature $H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}(t),\dot\gamma_{\mathcal{P}}(t))$ at $\gamma_{\mathcal{P}}(t)$ and the dual embedding curvatures $H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}(t),\dot\gamma_{\mathcal{D}}(t))$ at $\gamma_D(t)$, and the primal-dual curvature $h_{PD}(t)$.*

$$\|\frac{1}{2}H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}}(t),\dot\gamma_{\mathcal{P}}(t))\|^2_{\gamma_{\mathcal{P}}(t)} + \|\frac{1}{2}H_{\mathcal{D}}(\dot\gamma_{\mathcal{D}}(t),\dot\gamma_{\mathcal{D}}(t))\|^2_{\gamma_{\mathcal{D}}(t)} = h_{PD}(t)^2.$$

One of the important consequences of Theorems 6.3 and 6.4 is that the primal/dual path-following algorithm is better than the primal-dual path-following algorithm in view of the integral, and primal-or-dual algorithm is probably the best.

7 Concluding discussion

In this paper, we made an attempt to bridge the computational complexity and information geometry. We developed a suitable geometrical machinery for studying complexity of the polynomial-time interior-point algorithms. Upon this basis, there are several topics to be studied further. We discuss some of them below.

First topic we address is development of information geometric approach to primal-dual interior-point algorithms for symmetric cone programming. We expect that Euclidean Jordan algebra plays fundamental role here. In this paper, we established information geometrical meaning of the "primal-dual curvature" studied by Monteiro and Tsuchiya for classical linear programming. It would be interesting to introduce an analogue of the primal-dual curvature in the context of symmetric cone programming based on information geometry and develop the analogous iteration-complexity estimate based on the extended primal-dual curvature integral.

Second is the issue of the role of divergence in our analysis. In many application domains of information geometry, divergence plays important roles. In this paper, we did not touch on divergence at all. It would be interesting to look into more detail what is the role of divergence behind. We guess that the reason why we did not need divergence in our analysis is the part is implicitly taken care of by Theorems 3.1 and 3.2 which was developed by Nesterov and Nemirovski.

Third issue concerns possible application of geometry to development of robust implementation. Though semidefinite programming is efficiently solved with the interior point algorithms, some problem is solved yet difficult to solve, in particular those which arises in polynomial optimization. When the algorithm stacks, it is not clear whether the algorithm cannot make progress because of the nonlinear nature of the problem or because of rounding errors or not. The geometrical approach proposed here may shed new light on this problem.

Fourth issue is development of finer approximation or estimation of the number of iterations, through expansion by polynomials of β. In this paper, we approximated the number of iterations to trace the central path from $\gamma_{\mathcal{P}}(t_1)$ to $\gamma_{\mathcal{P}}(t_2)$ by the integral $\int_{t_1}^{t_2}(\|H^*_{\mathcal{P}}(\dot\gamma_{\mathcal{P}},\dot\gamma_{\mathcal{P}})\|^{1/2}/\sqrt{2})dt$ multiplied by $1/\sqrt{\beta}$. We may consider an expansion like

$$\text{(the number of iterations)} \sim \frac{p(\sqrt{\beta})}{\sqrt{\beta}},$$

where $p(\cdot)$ is a polynomial whose constant term is the integral and other coefficient probably has the form of integrals over the interval $[t_1,t_2]$. Like in statistics, we are not sure such expansion makes sense in practical applications, but may provide some deeper insight into the nature of interior-point algorithms and underlying information geometric structure of the optimization problems.

Finally, we mention possible application of self-concordance to other domains in mathematics. From the viewpoint of computational mathematics, self-concordance condition contains rich structures which enabled polynomial-time complexity analysis of the interior-point algorithms. Therefore, studying information geometry and/or the Hessian geometry induced by self-concordant barrier functions is interesting in its own right. Apart from the pure mathematical viewpoint, self-concordance may play significant role in other application areas of information geometry such as statistics and learning, signal processing, etc.

8 Appendix A: Proof of Proposition 3.4

In this appendix, we prove Propositions 3.4. To this end, we prove a weaker result that

$$\|\hat{y}-y\|_y \le \frac{16}{9}\eta. \tag{55}$$

We consider the Newton iteration initiated at y to compute a point $\hat{y}$ such that $g(\hat{y})=\hat{g}$ ($\hat{g}$ is given). Due to Theorem 3.2, the generated sequence is ensured to converge to $\hat{y}$. Let $\{y^k\}$ be the generated sequence ($y^0=y$), and let δ_k be the Newton decrement at y^k. By using $\|y^{k+1}-y^k\|_{y^k}=\|N_{y^k}(\hat{g})\|_{y^k}$, it is not difficult to see by induction and Theorem 3.2 that

$$\delta_k \le \left(\frac{16}{9}\right)^{2^k-1}(\eta)^{2^k} = \left(\frac{16\eta}{9}\right)^{2^k-1}\eta \le \left(\frac{1}{9}\right)^{2^k-1}\eta \le \left(\frac{1}{9}\right)^{k}\eta, \quad k=0,1,\ldots$$

Therefore, we have

$$\delta_0=\|y^1-y^0\|_{y^0}\le\eta,\ \|y^2-y^1\|_{y_1}\le\frac{1}{9}\eta,\ldots,\ \|y^{k+1}-y^k\|_{y_k}\le\left(\frac{1}{9}\right)^k\eta$$

(Note that $y^0=y$.)

Now we consider an auxiliary sequence $\{\alpha^k\}$ defined by the following recursion:

$$\alpha^0 = 0, \quad \alpha^1 = \eta \left(\le \frac{1}{16}\right), \quad \alpha^{k+1} := \alpha^k + \frac{1}{1-\alpha^k}\frac{1}{9^k}\eta, \quad k = 1, \dots$$

We claim that

$$\|y^k - y^0\|_{y^0} \le \alpha^k.$$

The proof is by induction. The claim holds obviously for $k = 0$ and 1. Suppose that the claim holds for $k \ge 1$. Then,

$$\|y^{k+1} - y^0\|_{y^0} \le \|y^{k+1} - y^k\|_{y^0} + \|y^k - y^0\|_{y^0} \le \alpha^k + \frac{1}{1-\alpha^k}\frac{\eta}{9^k} = \alpha^{k+1}.$$

Now, we introduce the sequence $\{\beta^k\}$ defined as follows:

$$\beta^0 = 0, \quad \beta^1 = \eta, \quad \beta^{k+1} = \beta^k + \frac{2\eta}{9^k} \quad k = 1, \dots$$

Then it is easy to see by induction that, if $\beta^k \le 1/2$, then $\alpha^k \le \beta^k$. Further more, we have

$$\beta^\infty = \eta + 2\eta\frac{1}{1-(1/9)}\frac{1}{9} = \eta(1+\frac{1}{4}) = \frac{5}{4}\eta.$$

Since $\eta \le 1/16$, we have $\beta^k \le \frac{1}{16}\cdot\frac{5}{4} < \frac{1}{2}$. This completes the proof of (55).

Now we prove (19). Let y_1 the point which we obtain after one Newton iteration. Since

$$\|y^2 - y^1\|_{y^1} \le \frac{16}{9}\eta^2,$$

we have, by using Theorem 3.1, Theorem 3.2 and (55),

$$\begin{aligned}\|y^\infty - y^0\|_{y^0} &\le \|y^\infty - y^1\|_{y^0} + \|y^1 - y^0\|_{y^0} \le \frac{1}{1-\eta}\|y^\infty - y^1\|_{y^1} + \eta \\ &\le \frac{1}{1-\eta}\frac{16}{9}\|y^2 - y^1\|_{y^1} + \eta \le \frac{1}{1-\eta}\left(\frac{16}{9}\right)^2\eta^2 + \eta \le \frac{4\eta^2}{1-\eta} + \eta.\end{aligned}$$

In the similar manner, we have

$$\eta - \frac{4\eta^2}{1-\eta} \le \|y^1 - y^0\|_{y^0} - \|y^\infty - y^1\|_{y^0} \le \|y^\infty - y^0\|_{y^0}$$

The bound (19) readily follows from this.

Now we prove the second claim.

If $\eta \le 1/16$, then we can apply the first statement, and then

$$\frac{\eta}{2} \le \zeta \le \frac{3\eta}{2},$$

and hence $\eta \le 2\zeta$. We also have

$$\frac{1}{1+8\eta} \le \frac{\eta}{\zeta} \le \frac{1}{1-8\eta}.$$

Therefore, we have

$$\frac{1}{1+16\zeta} \le \frac{\eta}{\zeta} \le \frac{1}{1-16\zeta}.$$

From this bound, we have, if $\zeta \le 1/64$, then,

$$1 - 22\zeta \le \frac{\eta}{\zeta} \le 1 + 22\zeta.$$

This completes the proof.

9 Appendix B: Proof of Theorem 4.3

In this appendix, we prove Theorem 4.3. As was mentioned before, to this end, we analyze $\{s^k/t^k\}$ rather than $\{s^k\}$ itself. We identify $\{s^k/t^k\}$, which the sequence on the dual feasible region $\mathcal{D}$, with a sequence generated by a polynomial-time path-following algorithm for the dual problem (2).

9.1 A predictor-corrector algorithm for the dual problem (26) and its polynomiality

First, we focus on (26) (or (2))

$$\min\ \langle d, s\rangle, \quad \text{s.t. } s \in \text{cl}(\mathcal{D}).$$

We rewrite this problem as an optimization problem with respect to y.

$$\min -\langle b, y\rangle \ \text{ s.t. } y \in \mathcal{T} \equiv \{y \mid c - A^{\mathsf{T}}y \in \Omega^*\}.$$

Let

$$\phi(y) = \psi^*(c - A^{\mathsf{T}}y).$$

Then, $\phi(y)$ becomes a self-concordant barrier with parameter p. The Legendre transformation is $g = -\partial\phi(y)/\partial y$. The central trajectory is written as the set of solutions to

$$-tb - g(y) = 0, \tag{56}$$

when t is changed from 0 to ∞. Let us denote by $y(t)$ the solution to (56). $y(t)$ is essentially the central trajectory for the original dual problem (26).

We define the neighborhood $y(t)$ of the central trajectory $y(t)$ as

$$\mathcal{N}_t^{\mathcal{D}}(\beta) = \{y \in \mathcal{T} \mid \|N_y(-tb)\|_y \leq \beta\}.$$

We will develop a predictor-corrector path-following scheme employing $-V_y^{\text{cen}} = -N_y(0)$ as predictor. We define

$$y_L(y, t, \Delta t) = y - \frac{\Delta t}{t + \Delta t}V_y^{\text{cen}} = y - \frac{\Delta t}{t + \Delta t}N_y(0).$$

We will consider an updating scheme

$$\begin{pmatrix} t \\ y \end{pmatrix} \Rightarrow \begin{pmatrix} t + \Delta t \\ y_L(y, t, \Delta t) \end{pmatrix}$$

at the predictor step. Nonlinearity of the stepsize function is to make match with the predictor-corrector algorithm in Section 4.2 which we will analyze in the next subsection.

Now we present a predictor-corrector algorithm for (26) based on this idea.

[Predictor-corrector algorithm for (26)]

1. Let $\beta \le 1/4$, and let $1/2 \le \eta$.

2. Let $y \in \mathcal{D}$ such that $y \in \mathcal{N}_t^{\mathcal{D}}(\frac{16}{9}\beta^2)$.

3. (Predictor step)

   Let $\Delta t > 0$ be such that

$$\|N_{y_L(y,t,\Delta t)}(-(t+\Delta t)b)\|_{y_L(y,t,\Delta t)} \in [\beta(1-\eta), \beta].$$

   Let $y^{\mathrm{P}} := y_L(y,t,\Delta t)$.

4. (Corrector step) At y^{P}, compute the Newton direction for the corrector-step $N_{y^{\mathrm{P}}}(-(t+\Delta t)b)$, and let $y^+ = y^{\mathrm{P}} + N_{y^{\mathrm{P}}}(-(t+\Delta t)b)$.

5. $t := t + \Delta t,\ y := y^+$ and return to step 1.

Now we have the following lemma on the admissible stepsize.

Lemma 9.1 *Let $0 < \beta_1 < \beta_2 < 1$, and let $y \in \mathcal{N}_t^{\mathcal{D}}(\beta_1)$. Let $\lambda(\beta_1, \beta_2)$ be the positive solution of the following equation:*

$$\frac{\left(1+\frac{\lambda}{\sqrt{p}}\right)\beta_1}{1-\lambda} + \frac{2\lambda^2}{(1-\lambda)^2} = \beta_2.$$

Then it is possible to take a step at least

$$\Delta t \ge t\frac{\lambda}{\sqrt{p}}$$

without violating the condition that $y - (\Delta t/(t+\Delta t))V_{\mathrm{cen}} \in \mathcal{N}_{t+\Delta t}^{\mathcal{D}}(\beta_2)$

Proof. Fig. 10 illustrates the situation in g-coordinate. Let y_0 be the current point y. By

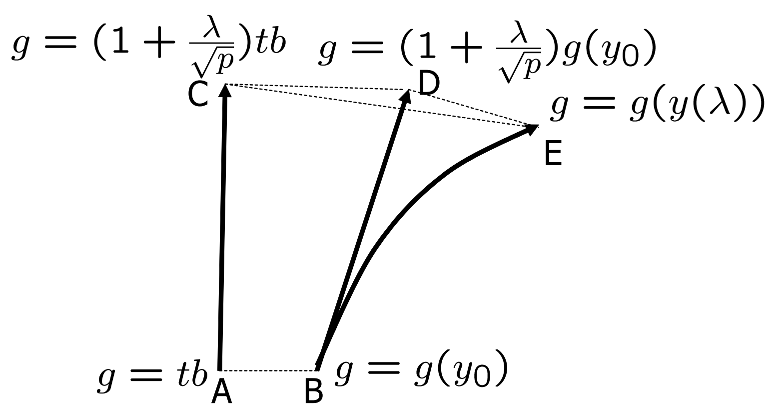

Figure 10: The predictor-corrector algorithm (in g-coordinate) .

assumption, we have $\|N_{y_0}(-tb)\|_{y_0} \le \beta_1$.

We introduce the step-size $\lambda = \sqrt{p}\frac{\Delta t}{t}$. Then, we have

$$\frac{\Delta t}{t+\Delta t} = \frac{\lambda}{\lambda+\sqrt{p}}.$$

Let

$$y(\lambda) = y_0 - \frac{\lambda}{\lambda+\sqrt{p}} N_{y_0}(0) = y + N_{y_0}\left(\left(1+\frac{\lambda}{\lambda+\sqrt{p}}\right) g(y_0)\right),$$

and let $g(\lambda) = g(y(\lambda))$ (the second equality follows from (18)). $g(y(\lambda))$ is a curve in g-coordinate though it is straight in y-coordinate.

In the following, we use the notation

$$\|u\|_y^- = \sqrt{u^\top G_y^{-1} u}.$$

For any $\hat{g} \in g(\mathcal{C})$, we have

$$\|N_y(\hat{g})\| = \|\hat{g} - g(y)\|_y^-.$$

We evaluate the Newton decrement for the point C at the point E in Fig. 10, i.e.,

$$\left\|N_{y(\lambda)}\left(\left(1+\frac{\lambda}{\sqrt{p}}\right) b\right)\right\|_{y(\lambda)} = \left\|\left(1+\frac{\lambda}{\sqrt{p}}\right) b - g(y(\lambda))\right\|_{y(\lambda)}^- .$$

For the purpose, we consider an intermediate point D, and we divide the norm into two parts, i.e.,

$$\left\|\left(1+\frac{\lambda}{\sqrt{p}}\right) b - g(y(\lambda))\right\|_{y(\lambda)}^- \le \left\|\left(1+\frac{\lambda}{\sqrt{p}}\right) b - \left(1+\frac{\lambda}{\sqrt{p}}\right) g(0)\right\|_{y(\lambda)}^- + \left\|\left(1+\frac{\lambda}{\sqrt{p}}\right) g(0) - g(\lambda)\right\|_{y(\lambda)}^- . \quad (57)$$

For later use, first we observe that the Newton decrement for the point E at the point B is bounded by λ. Due to (18) and Proposition 3.4, we have

$$\left\|\left(1+\frac{\lambda}{\lambda+\sqrt{p}}\right) g(0) - g(0)\right\|_{y_0}^- = \left\|N_y\left(\left(1+\frac{\lambda}{\lambda+\sqrt{p}}\right) g(0)\right)\right\|_{y_0} = \frac{\lambda}{\lambda+\sqrt{p}}\|N_{y(0)}(0)\|_{y(0)} \le \lambda < 1. \quad (58)$$

This means that $y(\lambda) \in D(y_0, \lambda)$.

Next we find a bound on the first term of (57), the norm of the displacement vector between C and D in Fig. 10, measured with the metric at E. We have the following bound

$$\left\|\left(1+\frac{\lambda}{\sqrt{p}}\right) tb - \left(1+\frac{\lambda}{\sqrt{p}}\right) g(0)\right\|_{y(\lambda)}^- = \left(1+\frac{\lambda}{\sqrt{p}}\right) \|tb - g(0)\|_{y(\lambda)}^- \le \frac{(1+\frac{\lambda}{\sqrt{p}})\beta_1}{1-\lambda}, \quad (59)$$

since $\|tb - g(0)\|_{y_0}^- \le \beta_1$ and $y(\lambda) \in D(y_0, \lambda)$.

Now we find a bound on the second term of (57). By add and subtract, we have

$$\begin{aligned}&\left\|\left(1+\frac{\lambda}{\sqrt{p}}\right) g(0) - g(\lambda)\right\|_{y(\lambda)}^- \\ &\quad\le \left\|\left(1+\frac{\lambda}{\sqrt{p}}\right) g(0) - \left(1+\frac{\lambda}{\lambda+\sqrt{p}}\right) g(0))\right\|_{y(\lambda)}^- + \left\|\left(1+\frac{\lambda}{\lambda+\sqrt{p}}\right) g(0) - g(y(\lambda))\right\|_{y(\lambda)}^- \quad (60)\end{aligned}$$

We have $\|g(0)\|^-_{y_0} = \|N_{y_0}(0)\|_{y_0} \le \sqrt{p}$ (Proposition 3.3) and $\|y(\lambda) - y_0\|_{y_0} \le \lambda$ ($y(\lambda) \in D(y_0, \lambda)$). Then, the first term of (60) is bounded as follows:

$$\left\| \left(1 + \frac{\lambda}{\sqrt{p}}\right) g(0) - \left(1 + \frac{\lambda}{\lambda + \sqrt{p}}\right) g(0) \right\|^-_{y(\lambda)} = \frac{\lambda^2}{\sqrt{p}(\lambda + \sqrt{p})} \|g(0)\|^-_{y(\lambda)}$$
$$\le \frac{\lambda^2}{\sqrt{p}(\lambda + \sqrt{p})(1-\lambda)} \|g(0)\|^-_{y_0} \le \frac{\lambda^2}{1-\lambda}. \tag{61}$$

On the other hand, it follows from (58) and Theorem 3.2 that

$$\left\| \left(1 + \frac{\lambda}{\lambda + \sqrt{p}}\right) g(y_0) - g(y(\lambda)) \right\|^-_{y(\lambda)} = \left\| N_y \left(\left(1 + \frac{\lambda}{\lambda + \sqrt{p}}\right) g(0) \right) \right\|_{y(\lambda)} \le \frac{\lambda^2}{(1-\lambda)^2}. \tag{62}$$

(We apply one iteration of the Newton method from y_0).

Compiling (57), (59), (60), (61) and (62), we obtain

$$\left\| N_{y(\lambda)} \left(\left(1 + \frac{\lambda}{\sqrt{p}}\right) b \right) \right\|_{y(\lambda)} \le \frac{\left(1 + \frac{\lambda}{\sqrt{p}}\right)\beta}{1-\lambda} + \frac{2\lambda^2}{(1-\lambda)^2}.$$

From this bound, the lemma immediately follows. ■

Theorem 9.2 *We have*

$$\Delta t \ge \frac{\lambda(\frac{16}{9}\beta^2, \beta(1-\eta))}{\sqrt{p}} t$$

at each iteration of the predictor-corrector algorithm, so that the algorithm generates the sequence y^k *such that*

$$y^k \in N_{t^k}(\beta), \quad t^k \ge \left(1 + \frac{\lambda(\frac{16}{9}\beta^2, \beta(1-\eta))}{\sqrt{p}}\right)^k.$$

Proof. The step Δt is chosen in such a way that

$$\beta(1-\eta) \le \|(t + \Delta t)b - g(y_L(y, t, \Delta t))\| \le \beta,$$

where $y \in \mathcal{N}^{\mathcal{P}}_t(\frac{16}{9}\beta^2)$. Letting $\beta_1 = \frac{16}{9}\beta^2$ and $\beta_2 = (1-\eta)\beta$. Then, taking account of $\beta \le 1/4$, we see that

$$\Delta t \ge \frac{\lambda(\frac{16}{9}\beta^2, (1-\eta)\beta)}{\sqrt{p}} t$$

yet satisfies the step-size condition above. ■

9.2 Polynomiality of the predictor-corrector algorithm

In the following, we will show that performing the predictor-step in [Implementable Algorithm] in Section 4.2 corresponds to moving in the negative centering direction in $\mathcal{D}$ when conic projection is performed. For $s \in t\mathcal{D}$, there exists unique y such that $s = tc - A^\top y$. We denote this y as $y(s)$, and let

$$u(s) \equiv \frac{y(s)}{t}.$$

Proposition 9.3 *Let $\bar{s} \in t\mathcal{D}$, $\bar{s} \in \mathcal{N}(\frac{16}{9}\beta^2)$ and perform the predictor step taking the step Δt. Then, we have*

$$u(\bar{s} + \Delta t\dot{\bar{s}}) - u(\bar{s}) = -\frac{\Delta t}{t + \Delta t} N_{u(\bar{s})}(0).$$

Proof. Since $s \in t\mathcal{D}$ and $\dot{s} \in \mathcal{D}$, we have

$$\begin{aligned}
\frac{s + \Delta t\dot{s}}{t + \Delta t} - \frac{s}{t} &= \frac{\Delta t}{t(t + \Delta t)}(t\dot{s} - s) \\
&= \frac{\Delta t}{t(t + \Delta t)}(-G_s^{-1}\Pi^{\perp}G_s s) \qquad \text{(use (36))} \\
&= -\frac{\Delta t}{t(t + \Delta t)} A^{\mathsf{T}}(AG_sA^{\mathsf{T}})^{-1}AG_s s \\
&= -\frac{\Delta t}{t(t + \Delta t)} A^{\mathsf{T}}(AG_sA^{\mathsf{T}})^{-1}Ax \qquad \text{(use (15) and } G_s = G_x^{-1}\text{))} \\
&= -\frac{\Delta t}{(t + \Delta t)} A^{\mathsf{T}}(AG_{s/t}A^{\mathsf{T}})^{-1}Ax(s/t) \qquad \text{(use (12) and (13))} \\
&= \frac{\Delta t}{(t + \Delta t)} A^{\mathsf{T}}(AG_{s/t}A^{\mathsf{T}})^{-1}A\nabla\psi^*(s/t) \\
&= \frac{\Delta t}{t + \Delta t} A^{\mathsf{T}}G_u^{-1}g(u) = -\frac{\Delta t}{t + \Delta t} A^{\mathsf{T}}N_{u(s)}(0).
\end{aligned}$$

■

This result implies that taking the step Δt in the predictor-corrector algorithm in Section 4.2 at a point $s \in t\mathcal{D}$ corresponds to taking $\Delta t/(t + \Delta t)$ along the negative centering direction.

One more proposition is needed to prove polynomiality.

Proposition 9.4 *Let $s \in t\mathcal{D}$. The following holds:*

1. $\|N_{u(s)}(-tb)\|_{u(s)} = \|N_s^{\mathbf{T}^*}(d)\|_s$.
2. $u(s) \in \mathcal{N}_t^{\mathcal{D}}(\beta)$ *iff* $s \in \mathcal{N}(\beta)$.

Proof. It follows from the discussion in the end of Section 3 that

$$N_{u(s)}(-tb) = (AG_{s/t}A^{\mathsf{T}})^{-1}(tb - Ax(s/t)).$$

This implies that, by using (12), (13) and (24),

$$\begin{aligned}
\|N_{u(s)}(-tb)\|_{u(s)} &= \sqrt{(tb - Ax(s/t))^{\mathsf{T}}(AG_{s/t}A^{\mathsf{T}})^{-1}(tb - Ax(s/t))} \\
&= \sqrt{(b - Ax(s))^{\mathsf{T}}(AG_sA^{\mathsf{T}})^{-1}(b - Ax(s))} \\
&= \|N_{y(s)}(-b)\|_{y(s)} = \|N_s^{\mathbf{T}^*}(d)\|_s.
\end{aligned}$$

■

Now we are ready to prove Theorem 4.3. Since

$$u(\bar{s}) \in \mathcal{N}_{t'}^{\mathcal{D}}(\beta) \Leftrightarrow \bar{s} \in \mathcal{N}(\beta) \cap t'\mathcal{D},$$

we see that the sequence $u(\bar{s}^k)$ is precisely regarded as a sequence generated by the predictor-corrector method for (26) analyzed in the previous section. Since we have

$$\frac{\Delta t^k}{t^k} \geq \frac{\lambda}{\sqrt{p}},$$

it is easy to see that the predictor-corrector algorithm generate a sequence satisfying

$$\bar{s}^k \in t^k\mathcal{D} \cap \mathcal{N}(\beta), \quad t^k \geq \left(1 + \frac{\lambda}{\sqrt{p}}\right)^k.$$

From this, Theorem 4.3 immediately follows.

References

[1] F. Alizadeh: Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, Vol. 5 (1995) pp. 13–51.

[2] F. Alizadeh and D. Goldfarb: Second-order cone programming. *Mathematical Programming*, Vol. 95 (2003), pp. 3–51.

[3] S. Amari: Differential geometry of curved exponential families — curvatures and information loss. *Annals of Statistics*, Vol.10 (1982), No. 2, pp. 357-385.

[4] S.-I. Amari: *Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics*, Vol. 28, Springer-Verlag, 1985.

[5] S. -I. Amari: Natural gradient works efficiently in learning. *Neural Computation*, Vol. 10 (1998), pp. 251–276.

[6] S.-I. Amari and H. Nagaoka: *Methods of Information Geometry.* AMS, Province, RI, USA. (An extended English translation of the book in Japanese published under the same title in 1993 from Iwanami-shoten, Tokyo, Japan.)

[7] E. R. Barnes: A Variation on Karmarker's algorithm for solving linear programming problems. *Mathematical Programming*, Vol. 36 (1986), pp. 174–182.

[8] D. Bayer, J. Lagarias: The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories. *Transactions of American Mathematical Society*, Vol. 314 (1989), No. 2, 499-526.

[9] D. Bayer, J. Lagarias: The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories. *Transactions of American Mathematical Society*, Vol. 314 (1989), No. 2, 527-581.

[10] A. Ben-Tal and A. Nemirovski: *Lectures on Modern Convex Optimization.* SIAM, Philadelphia, 2001.

[11] S. Boyd and L. Vandenberghe: *Convex Optimization. Cambridge University Press*, Cambridge, UK, 2004.

[12] G. B. Dantzig: *Linear Programming and Extensions.* Princeton University Press, Princeton, 1963.

[13] J.-P. Dediew, G. Malajovich and M. Shub: On the curvature of the central path of linear programming theory. Manuscript, arXiv:math/0312083, July, 2004.

[14] A. Deza, T. Terlaky and Y. Zinchenko: Central path curvature and iteration-complexity for redundant Klee-Minty cubes. *Advances in Mechanics and Mathematics*, Vol. 3 (2007), pp. 215–248.

[15] I. Dikin: Iterative solution of problems of linear and quadratic programming. *Soviet Mathematics Doklady*, Vol. 8(1967), 674–675.

[16] I. Dikin: On the convergence of an iterative process (in Russian). *Upravlyaemye Sistemi*, Vol. 12 (1974), pp. 54-60.

[17] J. J. Duistermaat: On the boundary behaviour of the Riemannian structure of a self-concordant barrier function. *Asymptotic Analysis*, Vol. 27 (2001), No. 1, 9–46.

[18] J. Faraut and A. Korányi: *Analysis on Symmetric Cones*, Cambridge University Press, Cambridge, UK, 1994.

[19] L. Faybusovich: Jordan algebras, symmetric cones and interior-point methods. Manuscript, University of Notre Dame, October 1995.

[20] L. Faybusovich: Jordan algebras and interior-point algorithms. *Positivity*, Vol. 1 (1997), pp. 331–357.

[21] L. Faybusovich: Linear systems in Jordan algebras and primal-dual interior-point algorithms. *Journal of Computational and Applied Mathematics*, Vol. 86 (1997), pp. 149–175.

[22] J. Gondzio and A. Grothey: Direct solution of linear systems of size 10^9 arising in optimization with interior point methods. In *Parallel Processing and Applied Mathematics PPAM 2005* (eds. R. Wyrzykowski, J. Dongarra, N. Meyer and J. Wasniewski), pp. 513–525, *Lecture Notes in Computer Science*, Vol. 3911, Springer-Verlag, Berlin, 2006.

[23] C. C. Gonzaga: Path-following methods for linear programming. *SIAM Review*, Vol. 34 (1992), pp. 167–224.

[24] C. C. Gonzaga and H. J. Lara: A note on properties of condition numbers. *Linear Algebra and its Applications.* Vol. 261 (1997), pp. 269–273.

[25] M. Iri: Integrability of vector and multivector fields associated with interior point methods for linear programming. *Mathematical Programming*, Vol. 52 (1991), pp. 511-525.

[26] F. Jarre: Interior-point methods for classes of convex programs. In *Interior-point Methods of Mathematical Programming* (ed. T. Terlaky), pp. 255–296, Kluwer, 1996.

[27] N. Karmarkar: A new polynomial-time algorithm for linear programming. *Combinatorica* Vol. 4 (1984), No. 4, pp.373–395.

[28] N. Karmarkar: Riemannian geometry underlying interior-point methods for linear programming. In *Mathematical Developments Arising From Linear Programming* (eds. J. C. Lagarias and M. J. Todd), pp. 51–75, *Contemporary Mathematics*, Vol. 114, AMS, Providence, 1990.

[29] M. Kojima, S. Mizuno and A. Yoshise: A primal-dual interior point algorithm for linear programming. In *Progress in Mathematical Programming, Interior-point and Related Methods* (ed. N. Megiddo), Springer-Verlag, New York, 1989, pp. 29–47.

[30] M. Kojima, S. Mizuno and A. Yoshise: A polynomial-time algorithm for a class of linear complementarity problems. *Mathematical Programming*, Vol. 44 (1989), pp. 1–26.

[31] L. McLinden: The complementarity problem for maximal monotone multifunction. In *Variational Inequalities and Complementarity Problems* (eds. R. W. Cottle, F. Giannessi, and J. L. Lions), pp. 251–270, John Wiley, New York, 1980.

[32] N. Megiddo: Pathways to the optimal set in linear programming. In *Progress in Mathematical Programming, Interior-point and Related Methods* (ed. N. Megiddo), pp. 131–158, Springer, New York-Berlin, 1989.

[33] N. Megiddo, S. Mizuno and T. Tsuchiya: A modified layered-step interior-point algorithm for linear programming. *Mathematical Programming*, Vol. 82 (1998), pp. 339-355.

[34] S. Mizuno, M. J. Todd, and Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, Vol. 18 (1993), pp. 964–981.

[35] R. D. C. Monteiro and I. Adler: Interior path following primal-dual algorithms. Part I: Linear programming. *Mathematical Programming*, Vol. 44 (1989), pp. 27–41.

[36] R. D. C. Monteiro and T. Tsuchiya: Polynomial convergence of primal-dual algorithms for the second-order cone program based on the MZ-family of directions. *Mathematical Programming*, Vol. 88 (2000), pp. 61–83.

[37] R. D. C. Monteiro and T. Tsuchiya: A variant of the Vavasis-Ye layered-step interior-point algorithm for linear programming. *SIAM Journal on Optimization*, Vol. 13 (2003), pp. 1054–1079.

[38] R. D. C. Monteiro and T. Tsuchiya: A new iteration-complexity bound for the MTY predictor-corrector algorithm. *SIAM Journal on Optimization*, Vol. 15 (2004), pp. 319–347.

[39] R. D. C. Monteiro and T. Tsuchiya: A strong bound on the integral of the central path curvature and its relationship with the iteration complexity of primal-dual path-following LP algorithms. *Mathematical Programming* (2007) (published on-line).

[40] K. Murota: *Discrete Convex Analysis*. SIAM, Philadelphia, 2003.

[41] H. Nagaoka and S.-I. Amari: Differential geometry of smooth families of probability distributions. *Mathematical Engineering Technical Reports*, Department of Mathematical Engineering and Instrumentation Physics, University of Tokyo, Tokyo, Japan, October 1982.

[42] Yu. Nesterov and A. Nemirovskii: *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, 1994.

[43] Yu. Nesterov and A. Nemirovski: Central path and Riemannian geometry. CORE Discussion Paper 2003/51, CORE, Belgium, July 2003.

[44] Yu. Nesterov and M. J. Todd: On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Foundations of Computational Mathematics*, Vol. 2 (2002), pp.333–361.

[45] A. Ohara: Information geometric analysis of semidefinite programming problems (in Japanese). *Proceedings of the Institute of Statistical Mathematics*, Vol. 46 (1998), no. 2, 317–334.

[46] A. Ohara: Information geometric analysis of an interior point method for semidefinite programming. In *Geometry in Present Day Science* (eds. O. E. Barndorff-Nielsen and E. B. V. Jensen), pp. 49–74, World Scientific, 1999.

[47] A. Ohara: Geodesics for dual connections and means on symmetric cones *Integral Equations and Operator Theory*, Vol.50 (2004), pp. 537–548.

[48] A. Ohara, N. Suda and S.-I. Amari: Dualistic differential geometry of positive definite matrices and its applications to related problems. *Linear Algebra and its Applications*, Vol. 247 (1996), pp. 31–53.

[49] J. Renegar: A polynomial-time algorithm, based on Newton's method, for linear programming. *Mathematical Programming*, Vol. 40 (1988), No. 1, pp. 59–93.

[50] J. Renegar: *A Mathematical View of Interior-point Methods in Convex Optimization. MPS-SIAM Series on Optimization*, SIAM, Philadelphia, 2001.

[51] H. Shima: *The Geometry of Hessian Structures*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2007.

[52] Gy. Sonnevend: An "analytical centre" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. *System Modeling and Optimization*, *Lecture Notes in Control and Information Sciences*, Vol. 84, pp. 866–875, Springer-Verlag, Berlin, 1986.

[53] G. Sonnevend, J. Stoer and G. Zhao: On the complexity of following the central path of linear programs by linear extrapolation: II. *Mathematical Programming*, Vol. 52 (1991), pp. 527–553.

[54] G. W. Stewart: On scaled projections and pseudoinverses. *Linear Algebra and its Applications*, Vol. 112 (1989) pp. 189–193.

[55] K. Tanabe: Geometric method in nonlinear programming. *Journal of Optimization Theory and Applications*, Vol. 30 (1980), pp. 181–210.

[56] K. Tanabe: Center flattening transformation and a centered Newton method for linear programming. (Manuscript presented at MP seminar, The Operations Research Society of Japan, July, 1987.)

[57] K. Tanabe: Centered Newton method for mathematical programming. In *System Modeling and Optimization* (eds. M. Iri and K. Yajima), pp. 197-206, Springer-Verlag, Berlin, 1988.

[58] K. Tanabe: Differential geometry of optimization (in Japanese). *Preliminary issue of the Bulletin of the Japan Society for Industrial and Applied Mathematics*, No. 3 (1990), pp.39–50.

[59] K. Tanabe: *Centered Newton Methods and Differential Geometry of Optimization. Cooperative Research Report*, Vol. 89, The Institute of Statistical Mathematics, November, 1996.

[60] K. Tanabe and T. Tsuchiya: New geometry of linear programming (in Japanese). *Suurikagaku*, No. 303 (1988), pp. 32–37.

[61] É. Tardos: A strongly polynomial algorithm to solve combinatorial linear programs. *Operations Research* Vol. 34, (1986) pp. 250–256.

[62] M. J. Todd: A Dantzig-Wolfe-like variant of Karmarkar's interior-point linear programming algorithm. *Operations Research*, Vol. 38 (1990) pp. 1006–1018.

[63] M. J. Todd, L. Tunçel and Y. Ye: Characterizations, bounds, and probabilistic analysis of two complexity measures for linear programming problems. *Mathematical Programming*, Vol. 90 (2001), No. 1, pp. 59–70.

[64] T. Tsuchiya: Global convergence of the affine scaling methods for degenerate linear programming problems. *Mathematical Programming*, Vol. 52 (1991), No. 3, pp. 377–404.

[65] T. Tsuchiya: Global convergence property of the affine scaling method for primal degenerate linear programming problems. *Mathematics of Operations Research*, Vol. 17, (1992) pp. 527–557.

[66] T. Tsuchiya: A convergence analysis of the scaling-invariant primal-dual path-following algorithms for second-order cone programming. *Optimization Methods and Software*, Vol. 11 and 12 (1999), pp. 141-182.

[67] T. Tsuchiya and M. Muramatsu: Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems. *SIAM Journal on Optimization*, Vol. 5 (1995), No. 3, pp. 525–551.

[68] L. Tunçel: Approximating the complexity measure of Vavasis-Ye algorithm is NP-hard. *Mathematical Programming*, Vol. 86 (1999), No. 1, pp. 219–223.

[69] K. Uohashi and A. Ohara: Jordan algebras and dual affine connections on symmetric cones. *Positivity*, Vol. 8 (2004), pp. 369–378.

[70] R. J. Vanderbei, M. S. Meketon, and B. F. Freedman: A modification of Karmarkar's linear programming algorithm. *Algorithmica*, Vol. 1 (1986), pp. 395–407.

[71] S. Vavasis and Y. Ye: A primal-dual accelerated interior point method whose running time depends only on A. *Mathematical Programming*, Vol. 74 (1996), pp. 79–120.

[72] H. Wolkowicz, R. Saigal and L. Vandenberghe: *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications.* Kluwer Academic Publishers, Boston, 2000.

[73] G. Zhao: On the relationship between the curvature integral and the complexity of pathfollowing methods in linear programming. *SIAM Journal on Optimization*, Vol. 6 (1996), pp. 57–73.

[74] G. Zhao and J. Stoer: Estimating the complexity of a class of path-following methods for solving linear programs by curvature integrals. *Applied Mathematics and Optimization*, Vol. 27 (1993), pp. 85–103.